Black Holes and Berry's Phase: Some aspects of geometry in physics.

Rodney A. Morgan

11 September 1995

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science at the University of Cape Town.

Abstract

The problem of the backreaction resulting from particle creation by black holes is examined from several vantage points. We first focus attention on the occurrence of the Berry phase in certain situations. This gives some insight into the geometry of quantum mechanics. Then we turn our attention to the analysis of quantum fields in black hole spacetimes. This brings us to the renormalisation of the stress- tensor via analytic methods. Finally, we draw on recent results to show how the Berry phase comes into play.

Contents

Introduction

Twenty years ago, Hawking [25] uncovered a remarkable relationship showing how quantum fields propagating in a black hole spacetime experience thermal effects. This discovery provided a major impetus for the further analysis of quantum field theory in curved spacetime. It points towards a close connection between thermodynamics, geometry and quantum theory. A continuing avenue of analysis relates to the question of how the fields affect the spacetime, in turn. At the present time, this 'backreaction effect' as it is termed, may be examined within a semiclassical framework. Such a framework is provided by a modification of the classical Einstein equation, $G_{\mu\nu} = T_{\mu\nu}$, whereby we replace the right hand side by a suitable quantum analogue such that

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \langle T_{\mu\nu} \rangle.$$

The major focus of this approach is the requirement that the quantum stress-tensor be suitably well behaved so as to be compatible with the classical geometry on the left hand side. This is a non-trivial expectation, the satisfaction of which will occupy a large proportion of our work here. It will be seen that $\langle T_{\mu\nu} \rangle$ has to undergo a certain amount of manipulation, in the form of *renormalisation*, in order to satisfy our requirements. We shall adopt a rigorous approach towards this goal. This entails a formulation of quantum field theory quite different from the usual. In this way, we avoid unnecessary attention to such details as a notion of 'particles', which ensures that we are wholly committed to the task of determining $\langle T_{\mu\nu} \rangle$.

If Hawking's result has shown how gravity affects things quantum, then the discovery by Berry [9] of a hitherto latent connection one-form in certain quantum systemshas certainly precipitated a somewhat wider perspective of how *geometry* affects quantum theory. We shall thus be adopting a somewhat less restrictive attitude in allowing for this possibility, *viz.* that gravity is not the only 'source' of geometric effects. Although the origins of Berry's discovery lie in molecular effects, it has stimulated myriad extensions [40] in quite esoteric directions, some of which we shall be reviewing here. Our choice of topics is not entirely aleotory, however. Despite the liberal view alluded to above concerning the role of geometry in quantum theory, we

have deliberately chosen to examine those implications of the Berry phase which have a definite 'gravitational' bent. Thus it will not come as too much of a surprise that our intention is to exhibit here a possible link between the Berry phase and the Hawking effect: specifically, we shall show that the backreaction mentioned above may be dealt with in certain cases using some of the techniques developed for the Berry phase [11]. Our approach throughout may thus be termed 'semiclassical', which provides a clue to the amalgamation we propose: in some cases, the Berry phase occurs in quantum systems dependent on parameters inhabiting some classical parameter space; in the study of the backreaction we exhibit, the (classical) gravitational field acts as parameter space for the quantum system in a minisuperspace model.

Our work is planned as follows: we introduce in 1.1 the notion of Hilbert space in a geometric fashion as a precursor to the derivation of the Berry phase in its original context. This introduction is fundamental to our work; herein we propose an original interpretation of the construction of Hilbert space in non-trivial spacetimes. This is followed by the above-mentioned derivation, along with its interpretation by Simon [41]. We then proceed to demonstrate some of the extensions and associated ideas of the original Berry phase. In section 1.3, we examine the proposals of [1],[2] concerning the realisation of a new state space for quantum mechanics. This shows the projective aspect of the geometry present in Hilbert space. We move on then to considerations of how the Berry phase fits in with the other geometric theory in modern physics, *viz.* gauge theory. We also consider an extension of the Berry phase in this context. Following this somewhat lengthy analysis, we turn our attention to symmetry in section 1.5. There we introduce the Born-Oppenheimer approximation, which we shall need for our promised *finalé*. We then focus on holonomy effects when parameter space is identified with (a globally hyperbolic) spacetime, following[23]. To conclude the discussion, we examine the notion of gravitational phase factor; hereafter (section 1.8), we engage in some speculation concerning possible future directions of research.

Having taken (temporary) leave of the Berry phase and quantum mechanics, we then examine field theory. Our uneasiness with the standard approach

is displayed in chapter 2, including a refutation of the notion of particles in section 2.1.1. But to criticise without suggesting an alternative would itself be questionable. We thus discuss a rigorous formulation of quantum field theory, using the theory of distributions [35],[54] in section 2.2. We utilise this formulation to discuss unitary equivalence, or what is known in conventional physics parlance as the S-matrix. This is useful for cases where a useful notion of particle does exist.

With this machinery in place, we tackle quantum fields in curved spacetime. This will require the geometric notions outlined in section 3.1. then we trun our attention shifts to the interaction between gravity and quantum fields known as the Hawking effect, *i.e.* particle creation by black holes. This is followed by a discussion of the backreaction problem, which takes up the rest of our work, wherein is shown the difficulties with the quantum stress-tensor. We outline the nature of renormalisation, using the ideas of the effective action. Thereafter, we assume a more practical stance using the technique of point splitting. Developing this method, however, is quite demanding: we introduce the algebraic notion of state and the restrictions thereon; this requires the development of some aspects of the Klein-Gordon field in globally hyperbolic spacetimes and the associated algebra and notion of state. This is finally brought to fruition in section 3.4.3, where we display the technique of obtaining $\langle T_{\mu\nu} \rangle$ with the concept of Hadamard states.

In conclusion, we complete the loop, as it were, by examining how the Berry phase may be of some value in the backreaction problem.

In the appendices, we exhibit the notation we have used and some concepts related to the algebraic approach.

Chapter 1

Berry Phase

1.1 Hilbert Space

Since it is our intention to examine the possible role of geometry in quantum theory, we use the machinery of differential geometry to examine the basic structures on which quantum mechanics rests, viz. the concepts of Hilbert space and operators on it, and the manner in which geometry comes into play. We shall be returning to this concept quite regularly, and thus exhibit what is hopefully a comprehensive outline.

We begin by recalling the standard construction of a Hilbert space[53]: we have in place a vector space, V, (over the field of complex numbers, $\mathbb{C}$), on which is defined a map $i : V \times V \to \mathbb{C}$ called the inner product; we denote this map by $i(v_1, v_2)$ or simply (v_1, v_2) for $v_1, v_2 \in V$, and require that it satisfy the following:

$$\begin{aligned} (v_1, v_2 + v_3) &= (v_1, v_2) + (v_2, v_3) \\ \overline{(v_1, v_2)} &= (v_2, v_1) \\ (v, v) &\geq 0 \end{aligned}$$

with equality holding for $v = 0$. V is thus an inner product space, with norm given by $\| v \| = \sqrt{(v, v)}$. If we have further that V is complete, *i.e.* all Cauchy sequences converge, then V has the structure of a Hilbert space,

denoted $\mathcal{H}$. Since all finite dimensional vector spaces are complete, they are clearly Hilbert spaces.

We clarify here some nomenclature which sometimes causes confusion: elements of $\mathcal{H}$ are referred to as state vectors, as they are members of a vector space. In the case of $\mathcal{H}$ being formed from a space of functions (as we shall demonstrate), the elements are also referred to as *wavefunctions*, for obvious reasons. We shall later describe how the states of quantum systems are described by elements of Hilbert space.

We define an operator L as a linear map $L : \mathcal{H} \to \mathcal{H}$. If L is bounded(i.e. there exists a $b \in I\!R^+$ such that $\| A\phi \| \leq b \| \phi \|$ for all$\phi \in \mathcal{H}$), we can define its adjoint $L^\dagger$ by the relation

$$(L^\dagger w, v) = (w, Lv) \tag{1.1}$$

If $L^\dagger = L$, then L is termed *self-adjoint* or Hermitian; if $L^\dagger = L^{-1}$, then L is said to be *unitary*. According to the Copenhagen interpretation of quantum mechanics, this distinction between operators is intimately related to the two types of processes which the wavefunction is supposed to undergo:

1. a continuous, linear, reversible, deterministic evolution based on the Schrödinger equation when no "measurement" is being made;

2. an apparently discontinuous, non-linear, irreversible and indeterministic evolution during a measurement.

The Hermitian operators are associated with the latter type of evolution; that this is so follows from the result that the set of physical observables (that is, those things that are 'measurable') is in one-to-one correspondence with the set of linear Hermitian operators on Hilbert space with complete orthonormal sets of eigenvectors.[1] Of course, the mechanism of measurement is an unexplained procedure, being defined only as a coupling between the quantum system and a suitable (classical) measuring apparatus.

We turn now to the construction of $\mathcal{H}$ in a differential geometry setting[7]: let $\mathcal{M} = M \times I\!R$ be a spacetime manifold, M being a connected orientable

[1]That not all Hermitian operators correspond to physical observables follows from the existence of *superselection rules*[43].

Riemannian manifold. Let $P(M,G)$ be the principal fibre bundle where the connection associated to the particle lives. Let $E(M,\alpha,\mathbb{C}^n)$ be the Hermitian bundle associated to an n-dimensional unitary representation α of G in the space $\mathbb{C}^n$ of internal degrees of freedom of the particle. Evolution will be considered to take place with respect to $t \in \mathbb{R}$. For each $t \geq 0$, $\pi_P^{-1}(M \times \{t\})$ is a principal fibre bundle with base M and structural group G, and $\pi_E^{-1}(M \times \{t\})$ is its associated vector bundle by means of α. Recall that a cross section σ is a map $\sigma : M \to B$, where B is an arbitrary bundle over M. Let $\mathcal{E}_t$ denote the set of sections ξ of $\pi_E^{-1}(M \times \{t\})$ such that

$$\int_M d\mu(\mathbf{x}) h(\xi(\mathbf{x}), \xi(\mathbf{x})) < \infty, \tag{1.2}$$

where $d\mu$ is the Riemannian measure of M and h indicates the product in the Hermitian structure of E. We postulate these as candidates for elements of $\mathcal{H}$, and define an inner product by

$$h(\xi,\eta) = \int_M h(\xi(\mathbf{x}), \eta(\mathbf{x})) d\mu(\mathbf{x}) \tag{1.3}$$

for any $\xi, \eta \in \mathcal{E}_t$. The need for a Riemannian measure is not essential. One could equally well use a Lebesgue measure[34]. We stress that it is the functions, σ, themselves which comprise the Hilbert space, rather than the function *values*. We are making here the elementary distinction between the function and its range; thus it is *not* E which forms the Hilbert space. (This is the promised construction of a Hilbert space as a function space.)

Before examining the geometric structures of operators, we point out here an original interpretation of how geometry non-trivially affects the quantum structure. First recall an elementary result from differential geometry [32]:

Theorem 1.1 *A principal bundle is trivial if and only if there exists a* global *section.*

The corresponding theorem for vector bundles follows from this result:

Corollary 1.2 *A vector bundle E is trivial if and only if its associated principal bundle $P(E)$ admits a global section.*

So local sections are, in effect, the expression of the non-trivial 'twisting' of a particular bundle. (In the situations we are particularly interested in, one may ascribe this twisting to gravitational effects.) Thus it is our contention that, in view of the fact that appropriate cross sections are interpreted as wavefunctions, one might interpret the local nature of the sections as a direct manifestation of the influence of geometry on quantum theory; geometrical considerations 'cause' the local nature of the sections. To what extent is this interpretation valid? We alluded above to the interpretation of the elements of Hilbert space as particle states; we shall show later how this view runs into severe difficulties when formulated in curved spacetime (i.e. we still have a Hilbert space structure in place, but the elements are no longer amenable to a particle interpretation). This justifies our opinion on the nature of local sections, in that global sections may be seen as particle states, but that local sections require a different view. Thus, a non-trivial geometric set-up naturally informs one of the limited extent of the particle interpretation. As far as we know, this is an original interpretation.

As for the geometric view of operators, we show in appendix B that the space of all bounded operators on a Hilbert space forms a Banach algebra, denoted $\mathcal{B}(\mathcal{H})$. (See also [7].) The set of bounded invertible operators with bounded inverse $\mathcal{L}(\mathcal{H}) = \{A \in B \mid A^{-1} \in \mathcal{B}(\mathcal{H})\}$ is an open Banach submanifold of $\mathcal{B}(\mathcal{H})$. Furthermore, $\mathcal{L}(\mathcal{H})$ endowed with the usual composition law, becomes an infinite dimensional Lie group. Its Lie algebra is $\mathcal{B}(\mathcal{H})$, where the Lie bracket is defined by $[A, B] = i(AB - BA)$. Now the set $\{\mathcal{A}_B\}$ of bounded self-adjoint operators in $\mathcal{H}$, $\mathcal{A}_B = \{\mathcal{A} \in \mathcal{B}(\mathcal{H}) | \mathcal{A}^\dagger = \mathcal{A}\}$, is a Lie subalgebra of $\mathcal{B}(\mathcal{H})$ which generates the unitary group $\mathcal{U}(\mathcal{H})$ through the exponential map. Therefore, $\mathcal{U}(\mathcal{H})$ is a closed Lie subgroup of $\mathcal{L}(\mathcal{H})$. Thus we can make the heuristic association:

$$\text{self-adjoint} \leftrightarrow \text{Lie algebra};$$
$$\text{unitary} \leftrightarrow \text{associated Lie group}.$$

In conclusion then, we have looked at here at what may be termed the 'kinematic' aspect of quantum theory, *viz.* the structure of the space of states

and the operators acting on it. In subsequent sections , we shall be revising our view of the former, as well as extending our reach from kinematics to dynamics.

1.2 Derivation and Interpretation

... the end of all our exploring
Will be to arrive where we started
And know the place for the first time. T. S. Eliot [15]

We turn our attention now to the promised consideration of the dynamics of quantum theory, i.e. we shall be examining here the description of evolution using the apparatus of the space of states and associated operators developed earlier. (Strictly speaking, we shall not be discussing fully-fledged dynamics, as such. Rather, we shall be looking at what may be called the border of dynamics and kinematics. This seemingly esoteric intent will become clearer as we proceed). Specifically, we shall be focussing attention on the particular aspect of geometry encountered in certain quantum systems first reported by Berry [9].

Berry considered a system governed by a Hamiltonian $\hat{H}$ dependent on varying parameters, $\vec{R} = (X_1, X_2, ...)$. [At this point, one may imagine $\vec{R}$ as a magnetic field, $\vec{B} = (B_x, B_y, B_z)$, say, which is time dependent, i.e. $\vec{B} \equiv \vec{B(t)}$, whereby the Hamiltonian acquires an implicit time dependence.] The evolution of this system between $t = 0$ and $t = T$ will be taken to be cyclic, i.e. it may be seen as transport round a closed path C in parameter space driven by the Hamiltonian, $\hat{H}(\vec{R}(t))$, with $\vec{R}(T) = \vec{R}(0)$. The state vector describing the system evolves via the Schrödinger equation

$$\hat{H}(\vec{R}(t))|\psi(t)\rangle = i\hbar|\dot{\psi}(t)\rangle. \tag{1.4}$$

If we prepare the system initially in one of the eigenstates of the Hamiltonian, $|n(\vec{R}(t))\rangle$ say, satisfying the eigenvalue equation

$$\hat{H}(\vec{R}) \mid n(\vec{R})\rangle = E_n(\vec{R}) \mid n(\vec{R})\rangle \tag{1.5}$$

with energy eigenvalue E_n, then the adiabatic theorem guarantees that if the evolution is 'slow enough', then the system will remain an eigenstate of the Hamiltonian, $|n(\vec{R}(t))\rangle$ at time t. Essentially then, the adiabatic theorem

implies that no energy level transitions may occur.[2] Thus $|\psi\rangle$ may be written as

$$| \psi(t)\rangle = e^{i\gamma_d} e^{i\gamma_B} |n(X(t))\rangle \tag{1.6}$$

where $e^{i\gamma_d}$ is the usual "dynamical" phase given by $\gamma_d = \int_0^t E_n(t)dt$. The second exponential is the object now referred to as the Berry phase: it is given by

$$\dot{\gamma}_n(t) = i\langle n(\vec{R}(t)) \mid \nabla_{\vec{R}} n(\vec{R}(t)\rangle \cdot \dot{\vec{R}}(t). \tag{1.7}$$

For a closed circuit C, the phase change is

$$\gamma_n(C) = i \oint_C \langle n(\vec{R}) \mid \nabla_{\vec{R}} n(\vec{R}\rangle \cdot d\vec{R} \tag{1.8}$$

the above result can be rewritten using Stokes' theorem whereby

$$\gamma_n(C) = -\int\int_C d\vec{S} \cdot \vec{V}_n(\vec{R}), \tag{1.9}$$

where $d\vec{S}$ is an area element in $\vec{R}$ space and

$$\vec{V}_n(\vec{R}) = Im \sum_{m \neq n} \frac{\langle n(\vec{R}) \mid \nabla_{\vec{R}} \hat{H}(\vec{R}) \mid m(\vec{R})\rangle \times < m(\vec{R} >}{(E_m(\vec{R}) - E_n(\vec{R}))^2}. \tag{1.10}$$

Now $\gamma_n(C)$ is independent of the choice of phase of the eigenstates $|n(\vec{R}(t))\rangle$; this means it is independent of the infinitely many Hamiltonians which may serve to carry the system round the circuit C, since to each phase we may associate an appropriate Hamiltonian. It is, in fact, the circuit C traversed in parameter space that is the source of the phase. The independence of $\gamma_n(C)$ of the choice of phase of the eigenstates means that it is *gauge invariant.* This means that the phase will be observable/measureable (see section 1.4.1).

Simon [41] showed that one may interpret the Berry phase as an holonomy: given x, we may construct a zero-energy eigenspace for $\hat{H}$, which acts as fibre to a line bundle over the parameter space, $\mathcal{M}$. This line bundle is embedded in the bundle $\mathcal{M} \times \mathcal{H}$, and inherits from this space a natural Hermitian

[2] Recall that, classically, 'adiabatic' means no heat loss. In the quantum case, this is translated to the requirement that no quanta are emitted.

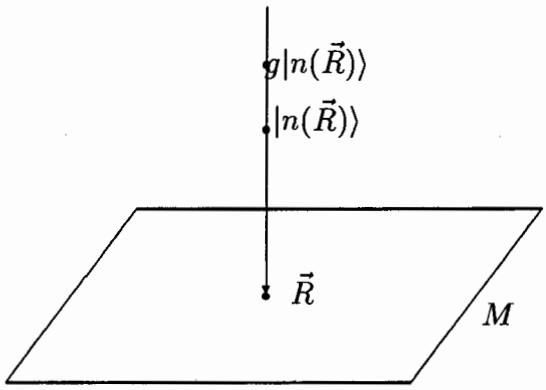

Figure 1.1: The fibre of a quantum mechanical system which depends on adiabatic parameters, $\vec{R}$.

connection. $\gamma_n(C)$ is then the integral of the curvature, and is given by

$$\gamma_n(C) = \int_S V \tag{1.11}$$

with $\partial S = C$ and

$$\begin{aligned} V &= i(d\phi, d\phi) \\ &= \sum_{i<j} \mathrm{Im}(\frac{\partial \phi}{\partial x^i}, \frac{\partial \phi}{\partial x^i}) dx_i \wedge dx_j \end{aligned} \tag{1.12}$$

The appearance of V here differs from the one in Berry and explicitly demonstrates that the phase is independent of the Hamiltonian.

To see how Berry's result relates to the classical concept of holonomy, we examine in a new way the familiar case of parallel transport on the sphere [40]. Imagine a unit vector $\hat{e}$ defined on the surface S^2. Let $\hat{r}$ be the unit radius vector; then the law of parallel transport on the sphere is embodied in the relation

$$\hat{e} \cdot \hat{r} = 0 \tag{1.13}$$

When we complete a closed circuit, C, on the sphere, we find that even though $\hat{r}$ returns to its original position, $\hat{e}$ does not. This effect is termed the holonomy, and is measured as the angle $\alpha(C)$ between the initial and final values of $\hat{e}$. To calculate $\alpha(C)$, we define $\hat{e}' = \hat{r} \wedge \hat{e}$ and

$$\hat{\psi} = (\hat{e} + i\hat{e}') \tag{1.14}$$

as a complex unit vector. Then the parallel transport law becomes

$$\text{Im } \psi^* \cdot d\psi = 0 \tag{1.15}$$

where $d\psi$ is the change in ψ resulting from a change $d\hat{r}$ ($*$ indicates complex conjugate). We can choose a local basis of unit vectors at each point, thus specifying a complex unit vector $\hat{n}(\hat{r})$, with which we may express $\hat{\psi}$ as

$$\psi = \hat{n}e^{-i\alpha} \tag{1.16}$$

Then since we require equation 1.15 to hold, we obtain for the holonomy

$$\begin{aligned} \alpha(C) = \oint d\alpha &= Im \oint \hat{n}^* \wedge d\hat{n} \\ &= \int\int_{\partial S=C} d\hat{n}^* \cdot d\hat{n} \end{aligned} \tag{1.17}$$

where Stokes theorem is used in the last over a surface S with C as boundary. A local change of basis at each point may be realised as a rotation $\mu(\hat{r})$ of the unit vector $\hat{n}(\hat{r})$ under which $\hat{n}(\hat{r}) \rightarrow \hat{n}(\hat{r})e^{i\mu(\hat{r})}$, leaving $d\hat{n}^* \cdot d\hat{n}$ invariant.

Heuristically, (as we are not here making claims towards any quantization of the model), the transition from classical transport to the quantum case involves replacing the complex unit vector ψ by a normalised quantum state $|\psi\rangle$ which is a unit vector in a Hilbert space, and substituting for $\hat{r}$ the parameters $X = (X_1, ...)$ of the system represented by $|\psi\rangle$. One may carry through the analogues of equations 1.13 to 1.17 to obtain the quantum version of holonomy. So far, this is just mathematics; we have merely been illustrating geometry without regard to the physics *per sé*. In particular, we have no means of obtaining a basis $|n\rangle$. It took the genius of Michael Berry to interpret the phase as in equations 1.5 to 1.9 described above.

1.3 Hilbert Space Revisited

Of the generalisations that followed in the wake of Berry's discovery, perhaps one of the most fruitful has been that made by Aharanov and Anandan [1]. We recall that Berry's main assumptions were the validity of the adiabatic approximation and the need for cyclic evolution. Aharanov and Anandan showed that the adiabatic assumption was sufficient but not necessary to obtain a (general) geometric phase, of which the Berry phase is a special example, by considering evolution in the projective Hilbert space $\mathcal{P}$ rather than in parameter space. $\mathcal{P}$ is defined via an equivalence relation, $\sim$ on the set of normalizable states in $\mathcal{H}$. This latter class, $\mathcal{N}$, is given by the relation

$$\mathcal{N} = \{|\psi> \in \mathcal{H}| < \psi|\psi> \neq 0\} \tag{1.18}$$

and $\mathcal{P}$ is obtained as $\mathcal{P} = \mathcal{N}/\sim$, where $\sim$ denotes that elements of $\mathcal{N}$ which differ only by a phase (i.e. a complex number) are regarded as equivalent. $\mathcal{P}$ is termed the set of rays [29]. Simon's view of the Berry phase as the holonomy in a line bundle over parameter space (where the cyclic evolution takes place) is here replaced by the notion that the evolution occurs in $\mathcal{P}$, with the phase residing in $\mathcal{N}$; thus the triple $(\mathcal{N}, \mathcal{P}, \pi)$ forms a principal fibre bundle with structure group $U(1)$, where π is the projection map taking a state in $\mathcal{N}$ to the ray on which it lies in $\mathcal{P}$.

The dispensable nature of the cyclic clause in Berry's derivation was illustrated by Samuel and Bhandari [37] who showed that for *any* path, it is possible to obtain a (generalised) geometric phase by closing it in a 'natural' manner. This 'tying up of loose ends' is attained by choosing a geodesic joining the two ends. A geodesic on $\mathcal{P}$ is the projection of a curve in $\mathcal{N}$ which satisfies the equation

$$\frac{D^2}{dl^2}|\varphi(l)> = \frac{d}{dl}|u> - iA_s|u> = 0. \tag{1.19}$$

where l is an affine parameter, $\varphi(l) > \in \mathcal{N}$ and $|u'>$ is the covariant derivative of $\varphi(l)$. This geodesic is, in turn, associated with a gauge invariant metric on $\mathcal{P}$ given by

$$ds^2 = < \partial_i n|(1 - |n>< n|)|\partial_j n > dX_i dX_j \tag{1.20}$$

The metric above, however is not positive definite, and does not determine a unique geodesic joining the two points, but it turns out that the geometric phase obtained is independent of the choice of geodesic. (One may regard the existence of this phase then as resulting from the introduction of a metric, which in the language of gauge theory, 'breaks the translational symmetry', thus allowing the introduction of a measurable quantity; see section 1.4.1). At this point it seems that not much more can be carried out. In the words of Shapere and Wilczek [40]: "It is hard to imagine anything more general than the geometric phase of Samuel and Bhandari, which applies to essentially any type of quantum evolution imaginable"!

Thus it comes as a major surprise that the story does not end here. We return to some notions from classical mechanics. Recall that dynamical evolution is taken to occur in a phase space: we specify a Hamiltonian, $H = H(q_1...q_n; p_1...p_n)$, as a function of the generalised coordinates $q_1...q_n$ and momenta $p_1...p_n$ which specify the states of a classical system; this H then determines evolution via Hamilton's equations

$$\frac{dq_\mu}{dt} = \frac{\partial H}{\partial p_\mu} \qquad \frac{dp_\mu}{dt} = -\frac{\partial H}{\partial q_\mu}. \tag{1.21}$$

We would like to move away from the notion of 'coordinates'; thus we let $y = (q_1...; ...p_n)$ and define a $2n \times 2n$ matrix, $\sigma^{\mu\nu}$, so that we can rewrite the above as

$$\frac{dy_\mu}{dt} = \Sigma_{\nu=1}^{2n} \sigma^{\mu\nu} \frac{\partial H}{\partial y^\nu}. \tag{1.22}$$

So far this has been just a formal construct. In keeping with our eventual aim of highlighting geometry in quantum theory, we re-interpret the states of the (classical) system as points in a $2n$- dimensional manifold, $\mathcal{M}$, on which is defined a fundamental structure, the symplectic form σ_{ab} (the inverse of the matrix above).

σ_{ab} is a type (0,2) tensor on $\mathcal{M}$ and $\sigma_{ab}v^b = 0$ iff $v^b = 0$ (i.e. σ_{ab} is nondegenerate). In this setting, the Hamiltonian, H, of the system is seen as a function on $\mathcal{M}$, $H : \mathcal{M} \to I\!R$. The Hamiltonian vector field, h^a on $\mathcal{M}$ is defined as

$$h^a = \sigma^{ab}\nabla_b H \tag{1.23}$$

The Hamiltonian equations of motion then correspond to the statement that the dynamically accessible regions of phase space are integral curves of h^a on $\mathcal{M}$. The utility of this construction is realised by the fact that the space of rays, $\mathcal{P}$ is, in fact, a $2n$ real dimensional phase space (when $\mathcal{H}$ has complex dimension $n+1$. The symplectic structure present occurs in addition to the projective geometry already resident in $\mathcal{P}$. The latter is the set of properties, such as collinearity, which are invariant under the group of non-singular linear transformations acting on $\mathcal{H}$. This collinearity property has the following physical implication: when we write

$$|\phi\rangle = c_1|\phi_1\rangle + c_2|\phi_2\rangle \tag{1.24}$$

we are, in fact, stating that the three rays to which $|\phi\rangle$, $|\phi_1\rangle$, and $|\phi_2\rangle$ belong are collinear points in $\mathcal{P}$ (with the coefficients c_1 and c_2 acting as coordinates of $|\phi\rangle$ in the coordinate system on the line). Now the above equation is a fundamental aspect of quantum theory, reflecting the notion of interference. Thus seen in this way, we have a geometrical aspect to quantum interference.

Furthermore, we may give a geometric interpretation to the inner product in $\mathcal{P}$: using the geometric phase (associated with the curve γ)

$$\beta = \oint_\gamma i < \tilde{\phi}|d\tilde{\phi} > \tag{1.25}$$

we may claim that $e^{i\beta}$, the holonomy transformation, gives a direct measure of this inner product. This is so because we have required that the inner product between neighbouring states have a zero imaginary part, i.e. $Im < \phi(s)|\phi(s+ds) >= 0$. But β is also the symplectic area of any surface spanned by γ with respect to σ. Using the coordinates (in some orthonormal basis) $Q^j = \tilde{\phi}_j$ and $P_j = i\tilde{\phi}_j^*$, we may write $\sigma = dP_j \wedge dQ^j$ as usual. Then the Poisson bracket between any two functions f, g on $\mathcal{P}$ can be written as

$$\{f,g\} = \frac{\partial f}{\partial Q^j}\frac{\partial g}{\partial P_j} - \frac{\partial f}{\partial P_j}\frac{\partial g}{\partial Q^j} \tag{1.26}$$

This brings us to the really interesting aspect of this work, which entails a drastic revision of quantum theory, even to the extent of re-examining the meaning of the wave function [2].

It all boils down to the concept of measurement. We wrote earlier (see section 1.1) that measurement, as opposed to Schrödinger-type eveolution, is conceived of as a discontinuous, non-linear, non-deterministic process. This ultimately results in the 'collapse of the wavefunction' scenario, the outcome of which can only be predicted on a probabilistic basis. So if we can control the measurement process such that collapse does not occur, we may well be able to refute the decades-old Copenhagen interpretation of quantum theory as a theory of statistical averages.

How does this occur? Precisely by the mechanism of *protective measurement.* This allows us to determine the expectation value of an observable for a wavefunction while it is prevented from collapse because of another interaction it undergoes at the same time (keeping in mind, of course, that a 'measurement' is really only a special sort of 'interaction'). Now this is where Berry's idea plays a major role: by preparing the system in an eigenstate of the Hamiltonian and making the measurement adiabatically, we ensure that no collapse occurs. This is essentially because we have no entanglement between the system and the apparatus, and thus no orthodox concept of measurement!

1.4 Gauge theories

We have seen that in addition to its intrinsic mathematical appeal, differential geometry provides a convenient framework for a variety of physical concepts. Now we shall demonstrate how the 'Standard Model' of modern particle physics is formulated in this language, albeit under the pseudonym of 'gauge theory'. We shall attempt a catalogue of the relevant structures which have hitherto been known as 'parallel transport' and 'holonomy', and how they apply to well-known physical theories. Our motivation for this study is the belief that in any attempt at examining 'foundational' structures, it would be wise to realise the role of gauge theories; this is true also for the Berry phase (and especially for what we would like to extend it to in future sections). The importance of gauge theory follows not only from its direct physical implications (as in the Standard Model), but also because it

purports to be a 'measurement theory', of sorts (being intimately involved in things 'measureable'). We will shortly see how this is implied. First, we examine the foundations of the theory, including the phenomenon of spontaneous symmetry breaking from a geometric standpoint. After laying these foundations, we show how this all ties in with the Berry phase; we also attempt to extend some of the ideas, following the work of[27]. Finally, we make some comments as to how these extensions tie-in with the work of[5].

1.4.1 Foundations

The first example of a gauge theory, though not formulated as such, is Maxwell's (classical) theory of electromagnetism. A (quantized) charged particle (of charge e) moving in a electromagnetic field is described by a wavefunction with a phase. This phase may be altered via an independent 'rotation' at every spacetime point without altering the values of measureable quantities. This requirement is termed 'local gauge invariance'. In fact, it may be taken as the *definition* of a measurable quantity, that it is invariant under gauge transformations and also that it commute with the gauge group.

To compare phases at different points, we introduce a set of spacetime dependent functions $a_\mu(x), \mu = 0, 1, 2, 3$. Then we say that the phase of a wavefunction, ψ, at x, is parallel to the phase at a neighbouring point $x + dx^\mu$ if the values of the phases differ by an amount equal to $e(a_\mu(x)dx^\mu)$. When we subject the phases to rotations, this implies a consistency condition on the a_μ. Specifically, we perform a so-called gauge transformation by rotating the phase of the wavefunction at x by an amount $e\alpha(x)$ depending on x, i.e.

$$\psi \rightarrow \psi'(x) = e^{ie\alpha(x)}\psi \tag{1.27}$$

Then the change of phase at $x + dx^\mu$ is given by

$$e\alpha(x) + \partial_\mu\alpha(x)dx^\mu \tag{1.28}$$

So to keep the phases 'parallel' as they were before the gauge transformation, we require that the $a_\mu(x)$ transform as follows:

$$a'_\mu(x) = a_\mu(x) + \partial_\mu\alpha(x) \tag{1.29}$$

This permits the parallel transport of wavefunctions from point to neighbouring point in spacetime, and by repeated application over finite distances along any path. Then by parallel transport from P to Q along path Γ, ψ acquires a change in the local value of the phase by an amount represented by $e\int_{P(\Gamma)}^{Q} a_\mu(x)dx^\mu$. However, this phase depends on the path Γ that we have chosen. For two paths Γ_1, Γ_2, we have

$$e\int_{\Gamma_1} a_\mu(x)dx^\mu - e\int_{\Gamma_2} a_\mu(x)dx^\mu = e\oint_{\Gamma_2-\Gamma_1} a_\mu(x)dx^\mu \tag{1.30}$$

By Stokes' thoerem, we can rewrite the RHS as a surface integral:

$$e\oint_{\Gamma_2-\Gamma_1} a_\mu(x)dx^\mu = -e\int\int_{\Sigma} f_{\mu\nu}(x)d\sigma^{\mu\nu} \tag{1.31}$$

over any surface Σ bounded by $\Gamma_2 - \Gamma_1$ with $f_{\mu\nu}(x) = \partial_\mu a_\nu(x) - \partial_\nu a_\mu(x)$. So if the parallel transport is to be path-independent, we require $f_{\mu\nu} = 0$, i.e. $a_\mu(x)$ must be 'curl-free'. Now this tensor $f_{\mu\nu}(x)$ is gauge invariant, whereby we mean that if we make the transformation 1.29 above, then

$$f_{\mu\nu}(x) \to f'_{\mu\nu}(x) = f_{\mu\nu}(x). \tag{1.32}$$

In physical terms, $a_\mu(x)$ is called the gauge potential, while $f_{\mu\nu}(x)$ is the (electromagnetic) field tensor. Mathematically of course they are the connection and curvature, respectively. (The transformations of the wavefunctions above indicate that the parallel transport of any particle (with some charge e) will be affected by these field, which are thus attributes of the spacetime under consideration (usually $\mathbb{R}^4$), rather than the particles themselves.) Thus, Maxwell's theory of electromagnetism may be formulated as a principal bundle over $\mathbb{R}^4$, with $U(1)$ as the structure group. A scalar field interacting with the Maxwell field is interpreted as a section of this bundle.

This brings us to an interesting point concerning the manner in which connections enter physics: the Lie algebra , $\mathcal{G}$, is isomorphic to the tangent space, T_eG, of a (not necessarily unique) Lie group G; the connection, being $\mathcal{G}$-valued, could be considered as the 'gradient' to some G-valued object; in particular, for 'physically interesting' cases, this analogy seems to hold. The

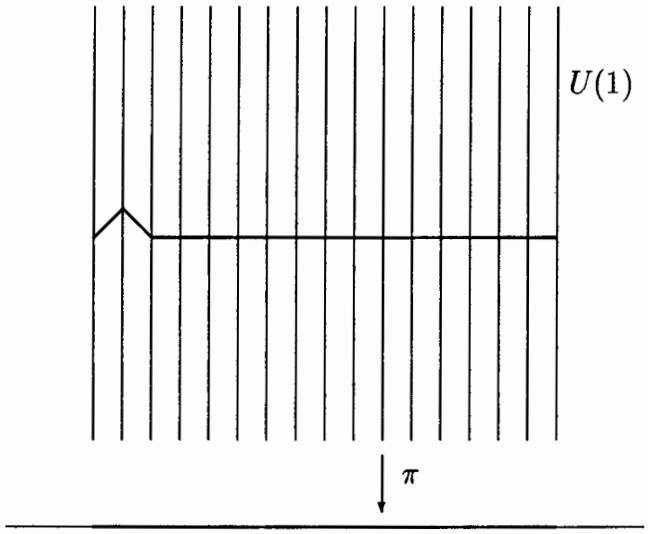

Figure 1.2: A scalar field may be seen as a section of a $U(1)$ fibre bundle over $\mathbb{R}^4$; π denotes the projection map.

traditional approach to the a_μ introduced above is to see it as the gradient of a potential $\varphi(x)$ i.e.

$$a_\mu(x) = \nabla\varphi(x) \qquad \mu = 1,2,3 \tag{1.33}$$

[with φ later interpreted as being the zeroth component of a_μ]. In general relativity, the analogy persists. There we have that the connection coefficients $\Gamma^{\alpha\alpha}_{\mu\nu}$ may be calculated using the components of the metric, g in the familiar manner [32], leading to the idea that the metric, g, may be considered as a potential of sorts for the $\Gamma^{\alpha}_{\mu\nu}$ in the same way that the φ was above, albeit in a more complicated form. We shall have more to say about the role of g in gauge theories subsequently.

To proceed with gauge theories proper, we examine briefly the suggestion made by Yang and Mills in 1954, that the way to generalise Maxwell's theory is to replace the $U(1)$ group by a more general one, possibly non-commutative. This met with initial resistance, but has since been verified and lauded as the foundation of all of physics. We shall not consider the im-

pressive successes which led to this acceptance, but allow ourselves to make a few remarks concerning it. If we look at the case of the group $SU(2)$ as was originally done by Yang and Mills we are led to discuss a two-component wavefunction $\psi = \psi^i(x), i = 1, 2$. By a change of phase here we shall mean a change in the orientation in internal space of ψ under a transformation $\psi \to S\psi$, where $S \in SU(2)$. Local gauge invariance now requires that the physics remains unaltered under $SU(2)$ transformations on ψ. We may proceed by analogy with the $U(1)$ case by introducing a set of functions $A_\mu(x)$, which allow the parallel transport of phases. These functions are required to be Hermitian and traceless, and thus are elements of the Lie algebra $\mathcal{SU}(2)$. Because of the non-Abelian nature of the group under discussion, one has to take particular care. Subjecting the phase to a transformation $S \in SU(2)$, we find the (famous) transformation law for the functions A_μ:

$$A_\mu(x) \to A'_\mu(x) = S(x)A_\mu(x)S^{-1}(x) - \frac{i}{g}\frac{\partial S}{\partial x^\mu}S^{-1}(x) \tag{1.34}$$

For S infinitesimal, i.e. $S(x) \approx (1 + ig\Lambda(x))$, we have:

$$A'_\mu(x) = A_\mu(x) + \partial_\mu\Lambda(x) + ig[\Lambda(x), A_\mu(x)] \tag{1.35}$$

to leading order in Λ. We also obtain a field tensor, $F_{\mu\nu}$, defined as

$$F_{\mu\nu}(x) = \partial_\nu A_\mu(x) - \partial_\mu A_\nu(x) + ig[A_\mu(x), A_\nu(x)] \tag{1.36}$$

Under a gauge transformation $\psi \to S(x)\psi(x)$, we have that $F_\mu(x)$ transforms covariantly as follows:

$$F_{\mu\nu}(x) \to F'_{\mu\nu}(x) = S(x)F_{\mu\nu}(x)S^{-1}(x). \tag{1.37}$$

This *covariant* transformation of $F_{\mu\nu}(x)$ is in contrast to the *invariant* transformation of the abelian $f_{\mu\nu}$ in 1.32 above. Essentially, this covariance is due to the fact that we are now dealing with a higher dimensional group, i.e. $SU(2)$ as opposed to $U(1)$. Heuristically, we are able to 'rotate' the field tensor values, $F_{\mu\nu}$, without changing the physics because these values now lie in a plane, of sorts (the $\mathcal{SU}(2)$ Lie algebra). This sense of rotation may be

seen from the transformation 1.37 above, the form of which is reminiscent of a familiar rotation in three-dimensional (real) space.

Experimentally, what is actually measured in experiments are the so-called Dirac phase factors, given by

$$\Phi(C) = Tr(e^{ig\oint_C A_\mu(x)dx^\mu}) \tag{1.38}$$

for the $SU(2)$ case, where Tr indicates the trace and

$$\Phi(C) = e^{ie\oint_C a_\mu(x)dx^\mu} \tag{1.39}$$

for the $U(1)$ case, where C is a closed curve. Both these quantities are gauge invariant. The reason why they are suitable for describing particular physical situations is due to the difficulties in dealing directly with the A_μ/a_μ or the $F_{\mu\nu}/f_{\mu\nu}$ as variables: the A_μ may be replaced as in 1.34 above without altering the phase difference, and so tend to overdescribe the system; while on the other hand, the $F_{\mu\nu}$ underdescribe the system, since $F_{\mu\nu}(x)$ specifies parallel transport only over infinitesimal loops, while $\Phi(C)$ is a global version of $F_{\mu\nu}$, which describes also loops of finite size. So where finite loops may be built up from infinitesimal ones, the descriptions given by $F_{\mu\nu}$ and $\Phi(C)$ would be equivalent. But there are cases where this is not so, such as the famous Aharanov-Bohm effect.

Up till now, we have illustrated the necessary mathematical machinery to deal with three of the known forces, viz the strong, weak and electromagnetic: the 'Standard Model' of physics describes these forces using the groups $SU(3), SU(2)$ and $U(1)$. It would be anomalous if we did not describe how two of these forces were unified using the ubiquitous phenomenon of spontaneous symmetry breaking. First we need the notion of a *Higgs field*: we shall see in section 2.1.1 that particles are characterised by their transformation properties under certain irreducible representations of the Poincaré group; thus we may associate particles with a representation $\varrho : G \to GL(V)$ of the gauge group, G, in some group of operators, $GL(V)$, over an appropriate vector space, V. (This V is, in fact the fibre of an associated vector bundle, E to the principal bundle, $P(M, G)$). Then we say that a particle of type ϱ interacting with the gauge field is described by a map $\varphi : P \to V$. The

pull-back of φ by a section s is called a Higgs field:

$$\Phi = \varphi \cdot s : \quad N \to V. \tag{1.40}$$

Then consider a Higgs field whose range is an orbit, W, of G in V, i.e. $\varphi : P \to W \subset V$, where W is such that for any $w_0, w \in W$, there exists $a \in G$ with $w = \varrho(a)w_0$. Let H be the isotropy group of w_0: $H = \{a \in G | \varrho(a)w_0 = w_0\}$. Then we have that $Q = \{p \in P | \varphi(p) = w\}$ is a sub-bundle of P over M with structure group H. The symmetry is said to be broken. On the other hand, given a reduction Q of P to $H \subset G$, we can define a Higgs field $\varphi : P \to W = G/H$ by putting $\varphi(p) = H \subset W$ for $p \in Q$. Finally, a connection form ω on P, restricted to Q, defines an H- connection on $Q \iff D\omega = 0$.

1.4.2 Including gravity

At this point it is well worth reminding ourselves that *gauging* a theory is a far cry from *quantizing* it. Nowhere is this more evident than in the theory of general relativity, where there has been considerable progress describing it as a gauge theory, but for which a suitable quantization procedure is still outstanding. It would be worthwhile to outline here the extent to which we may model gravity as a gauge theory, thus bringing it into line with the other forces, which brings with it some hope of a quantization scheme. As outlined above, the major aspects of describing forces using gauge theory is to choose a suitable gauge group. We identify $Diff(\mathcal{M})$ as the gauge group of gravity [36]. Some difficulties with gravity arise from its dissimilarity with Yang-Mills type gauge theories. In particular, the action for General Relativity is based on a Lagrangian linear in curvature, i.e. $\mathcal{L}_{GR} = \int d^4x R(x)\sqrt{-g(x)}$, while Yang-Mills type Lagrangians are generally quadratic in curvature, being of the form $\mathcal{L}_{YM} = \int d^4x F^{\mu\nu}F_{\mu\nu}\sqrt{-g(x)}$. Thus there exists in Yang-Mills type actions the gauge freedom associated with the tensor, $F_{\mu\nu}$, while for general relativity there is no such freedom since we are dealing with a scalar in the action, viz the Ricci scalar R. Also, in General Relativity, there exists a natural soldering of the bundle of linear frames, LM, to the underlying

manifold, M (this soldering may be seen as a manifestation of the Principle of Equivalence)[44]. This soldering relates the tangent space of the bundle to the tangent space on the base via a one-form $\vartheta : T(LM) \to I\!R^n$. Suppose $e = (e_\mu) \in LM, u \in T_e(LM)$, then $\vartheta^\mu(u) = \{T_e\pi(u)\}^\mu$ is the μ-th component of u projected onto M, with respect to the basis e on M.

If $s : N \to LM$ is a local section (i.e. a field of frames), then

$$s^*\vartheta^\mu = s^\mu \tag{1.41}$$

This soldering is the source of the difference between the Lagrangians mentioned above: using it, we construct various invariants which may act as the kinetic term in the gravitational Lagrangian, which are not available to theories without soldering. The soldering form is also a kind of Higgs field (see section 1.7); it differs from other Higgs fields in being a one-form (rather than a zero-form / function).

Before leaving this section, we point out that the whole aim, in a sense, of gauge theories is to identify the nature of observable quantities. In fact, observables are those quantities which obey the principle of gauge invariance stated earlier (and furthermore commute with the action of the gauge group). So to deal effectively with gravity as a gauge theory, we have to identify suitable candidates for labelling as 'observables'. This is a major snag [36]. The most 'obvious' candidate suitable for identification as an observable is the Ricci scalar, $R(x) = g_{\mu\nu}(x)R^{\mu\nu}(x)$. However, that this is not suitable follows from looking at its transformation properties under $Diff(\mathcal{M})$: for $f \in Diff(\mathcal{M})$, we have

$$\delta_f R(x) = f^\mu \partial_\mu R(x) \neq 0 \tag{1.42}$$

[These problems persist even at a (pseudo-) quantum level: assume that there exists a unitary representation of $Diff(\mathcal{M})$, so $f \to U(f)$. Then $U(f)R(x)U(f)^{-1} = R(f(x))$, illustrating that $R(x)$ does not commute with the gauge group, and is thus not suitable as an observable.] In passing we mention that when we encounter quantum fields, the following should be kept in mind: the difference between the gravitational field and classical fields which can be straightforwardly quantised lies in the fact that for a single

particle, there is a gravitational field description; but other fields generally involve infinitely many particles. This remark remains valid despite the later intrusion of gravitons into the picture: gravitons are particles belonging to a (partly) quantised gravitational field; our remark applies to the classical case.

1.4.3 Gauge theories and the Berry Phase

As we have shown above, ordinary gauge theories contain the following elements: (i) a base manifold, usually taken to be Minkowski spacetime $(\mathbb{R}^4, \eta_{ab})$, and (ii) an internal symmetry space, identified with a gauge group, such as $U(1)$. Thus, in a differential geometric setting, we are dealing with a $U(1)$ bundle over spacetime. We may seek to elaborate this structure by:

1. replacing the fibre, $U(1)$, by some non-abelian structure, such as $SU(2)$, which was the idea propounded by Yang and Mills;

2. replacing the base space $(\mathbb{R}^4, \eta_{ab})$

The previous section showed the result of the first elaboration; here we shall opt for the latter approach, following [27]. The first realisation is that it is possible to accomplish the stated goal in two ways. One way is to replace $(\mathbb{R}^4, \eta_{ab})$ by a curved manifold, denoted $(\mathcal{M}, g)$, whereby we are led to the Einstein-Maxwell theory, which requires a reconception of the very notion of spacetime, but does not further our aims of understanding gauge theories, as such. Instead, we seek the rewards which follow from the realisation that gauge theories constructed over $\mathbb{R}^4$ are degenerate in that $\mathbb{R}^4$ is identified with its own translation group! This means that any point in $\mathbb{R}^4$ can be obtained by applying a four-dimensional translation to the origin, and conversely, any four-dimensional applied to the origin specifies a point in $\mathbb{R}^4$. So we might as well consider a gauge theory over $\mathbb{R}^4$ as a gauge theory over the translation group manifold, which is abelian in this case. Then it is worthwhile to investigate what happens when we replace the abelain translation group by a non-abelian one, and formulate a gauge theory over *its* manifold.

Thus we are maintaining the spirit of the original generalisation of Yang and Mills, except that we shall now be dealing with the base, rather than the group. [Needless to say, since we shall require the underlying space to be a manifold, we necessarily consider only Lie groups.] It will be seen that we may recover the Berry phase in this way, as well as opening up other exciting possibilities.

Two phases at neighbouring points x^μ and $x^\mu + dx^\mu$ are said to be parallel if they differ (locally) by an amount $gA_\mu(x)dx^\mu$. Thus we interpret the gauge potential A_μ as giving us parallel transport of phases from point to point in the base manifold. When we replace the translation group of $I\!R^4$ by a non-abelian group, we are regarding the base space as a group so that each point in it corresponds to a group element; a displacement from one point to another may be affected by an action, such as left multiplication, by some element of the group. For a displacement between neighbouring points, we may use the action of a group element differing infinitesimally from the identity; in other words, we may use a member of the Lie algebra. As for the gauge potential, we may regard it as giving us parallel transport not from 'point to point' as above, but as a prescription for parallel transport of phases under the action of a non- commutative displacement algebra, which generalises the concept of translations over $I\!R^4$. (To keep track , we shall refer to this 'new' gauge potential as A_l.) Letting d_l denote the generators of this algebra, we have the commutation relations:

$$[dl, d_k] = C^k_{lm} d_k \tag{1.43}$$

A point ξ on the base may be displaced to a neighbouring point ξ', by acting on the left with $(1 + \epsilon^l d_l)$, for infinitesimal ϵ^l. Correspondingly, the wave-function changes to

$$\psi(\xi') = \psi(\xi) + \epsilon^l d_l \psi(\xi) \tag{1.44}$$

The gauge potential A_l will give us parallel transport under gauge transformations.

As for the field tensor, $F_{\mu\nu}$, we must remember that we have torsion in the base, i.e. a displacement d_l followed by d_k does not necessarily bring us to the same point as first bringing about a displacement d_k followed by d_l.

This modification results in a new formula law for the field tensor, viz

$$F_{lk} = [d_k, A_l] - [d_l, A_k] + ig[A_l, A_k] + C_{lk}^m A_m \tag{1.45}$$

By highlighting the group as base as we have attempted to illustrate here, we have arrived at a dual understanding of the following concepts:
(A) gauge invariance can be understood either as invariance under independent phase rotations at different points in the base space, or as invariance under phase rotations before and after the action of elements of the displacement algebra;
(B) the gauge potential can be regarded as giving us parallel transport of phases from point to neighbouring point in the base manifold, or as parallel transport of phases under the action of the displacement algebra.

The dual interpretation has so far helped to diplay the richness involved in this interpretation of gauge theories. The physical application of this arises, for example, when we see how Berry's phase is recovered by considering an atom with a deformed, heavy nucleus. The orientation of the latter is specified by Euler angles ε^μ, $\mu = 1, 2, 3$, or equivalently, by an element of the three- dimensional rotation group. Then the whole system (of nucleus and surrounding cloud of electrons) is described by a Hamiltonian, $H(\varepsilon)$, which is dependent on $\varepsilon \in SO(3)$, but acts on the electronic degrees of freedom. Theses are obtained by considering solutions to the eigenvalue equation

$$H(\varepsilon)|n;\varepsilon> = E_n|n;\varepsilon> . \tag{1.46}$$

The eigenstates, as before, have a phase freedom. Thus to each $\varepsilon \in SO(3)$ there is a circle representing the values the phase can assume, which leads toa $U(1)$ bundle over $SO(3)$. Then by consideing an adiabatic rotation of the nucleus by $\varepsilon(t)$ in time t, we may recover the Berry phase as

$$\gamma_B(t) = \int_0^t d\varepsilon^\mu < n;\varepsilon|\frac{\partial}{\partial\varepsilon^\mu}|n;\varepsilon> . \tag{1.47}$$

1.4.4 Duality...

We turn our attention now to a further extension of these results. As illucidated in the previous section, we view the wavefunction, ψ, as a map assigning

a phase to each spacetime point, and the gauge potential, A, as providing parallel transport of the phase from point to point. However, the fact that we were dealing with gauge structures over $\mathbb{R}^4$ allowed us the degeneracy luxury elaborated above, wherein both ψ and A_l admit an alternative interpretation.

Thus, ψ can also be thought of as a prescription of how its value changes under translations, and A_l as a parallel transport of phase under translations. We would now like to examine the implications of keeping for ψ the sole interpretation of being a prescription for how its value changes under the action of the d_l's. In this sense, we attach a more 'bootstrapping' interpretation of the wavefunction. Recall that the value of ψ at a point ξ', reached by the action of an infinitesimal displacement acting on ξ was, according to refeq:dis, represented by the action of the operator

$$(1 + \epsilon^l d_l) \tag{1.48}$$

operating on ψ (evaluated at ξ). So the operators d_l play the important role of telling us how the wavefunction changes under displacement; all we need then is to have ψ evaluated at a single point, ϵ say. then we may determine its value at any other point by the action of an appropriate $(1+\epsilon^l d_l)$ term acting on it. After that, we need no longer refer to points in the base space. The same considerations apply to the A_l too, which are also operators on the ψ. Thus we may parallel transport phases once we have determined the phase at a single point without referring again to the points in the base space. So we arrive at the conclusion that the information provided by the ψ and A in the case of the base space of points may be encoded in the operators d_l and A_l. We shall now investigate whether it is possible to construct a meaningful 'local gauge theory' with these operators, discarding the ψ and A for now, and with them the notion of points (in the base space).

What would be the use of this? As pointed out by the authors, in electrodynamics, for example, there already exits a valid conception of space and time, as well as the meaning of the field values at the points in the spacetime; this is well-known and perfectly acceptable. However, it may happen that under certain circumstances we may have to abandon these cherished notions. One is aware, in particular, that in quantum gravity one should

have to re-examine one's views on these matters. In particular, recent work by Anandan [5] shows the explicit use of this scenario: using the 'Hole Argument' of Einstein he concludes that "...in quantum gravity space-time points have no invariant meaning....Consequently, the space-time manifold, which appears to be redundant, may be discarded...". (We shall return to this point in section 1.7.)

However, here we are stressing that in keeping with the idea of a new view of quantum theory affected by geometric concepts, we consider the results of this extension in their own right. Thus we maintain the minimal structure necessary to construct a gauge theory. To this end, we start out with a Hilbert space, $\mathcal{H}$, to which the wavefunction, ψ, belongs. We postulate the existence of two algebras, $\mathcal{B}$ and $\mathcal{U}$ which act on $\mathcal{H}$. The former defines what is meant by displacements on ψ, and is generated by the d_l. We want $\mathcal{U}$ to contain all elements associated with the gauge structure; in particular, it should contain the gauge potential, A_l (which provides parallel transport). We then follow through the main steps shown earlier [in particular eqns. 1.29 to 1.36], but now with the new interpretation: under an infinitesimal displacement $(1 + \epsilon^l d_l)$, we have that the wavefunction, $\psi \in \mathcal{H}$, changes to a new element in $\mathcal{H}$, *viz.*

$$\psi \to (1 + \epsilon^l d_l)\psi = \psi + \epsilon^l d_l \tag{1.49}$$

As for A_l, we see that under the above displacement, parallel transport leads to another element given by

$$\psi \to \psi + ig\epsilon^l A_l \psi. \tag{1.50}$$

To deal with gauge transformations in this framework, we recall that a local gauge transformation is parametrised by $\Lambda(\xi)$, where ξ is a point in the base space, and Λ is a function taking its values in the gauge algebra. Now without any recourse to the base space, how are we to proceed? First recognise that Λ is ξ-dependent, and thus need not commute with elements of $\mathcal{B}$. We postulate then that the action of a gauge transformation on ψ is as follows

$$\psi \to \psi' = (1 + ig\Lambda))\psi \tag{1.51}$$

To see how the concepts of (translational) displacement, parallel transport and gauge transformations interrelate, consider what happens when we make a gauge transformation prior to performing a displacement:

$$\psi \to (1+\epsilon^l d_l)(1+ig\Lambda)\psi = (1+\epsilon^l d_l)(1+ig\Lambda)(1-\epsilon^l d_l)\tilde{\psi} \qquad (1.52)$$

where $\tilde{\psi} = (1+\epsilon^l d_l)\psi$ is the result of the displacement before the gauge transformation. This clearly shows the non- commutativity of the relationship between $\mathcal{U}$ and $\mathcal{B}$.

We defined ψ' in 1.50 above as being the parallel transport of ψ under displacement given by 1.48; with the presence of the gauge transformation, however, the displaced wavefunction is given by 1.51 above. So we should prescribe the parallel transport of $\psi'[= (1+ig\Lambda)\psi]$ under the displacement in eqn. 1.48 as given by $(1+\epsilon^l d_l)(1+ig\Lambda)(1-\epsilon^l d_l)(1+ig\epsilon^l A_l)\psi$.
From the equality

$$(1+\epsilon^l d_l)(1+ig\Lambda)(1-\epsilon^l d_l)(1+ig\epsilon^l A_l)\psi = (1+igA'_l)(1+ig\Lambda)\psi \qquad (1.53)$$

we may derive the transformation law for the gauge potential:

$$A'_l = A_l + [d_l, \Lambda] + ig[\Lambda, A_l] \qquad (1.54)$$

This is strikingly familiar, and gives one hope that we have not strayed too far afield. Indeed, not only does the gauge potential transform in the 'usual' manner, but we may also derive similar formulae to describe covariant derivatives and curvature, provided, of course, that we keep in mind the accompanying alterations in interpretation. For example, the covariant derivative, defined as $D_l = d_l - igA_l$, is now thought of as the difference between the value (of the wavefunction) obtained by parallel transport and that obtained under displacement; we are no longer in a position to regard it as giving the difference between the values at the displaced point (as comfortably familiar as this may seem). Similar remarks apply to the field tensor F_{lk}, which remains a covariant curvature, but is stripped of its interpretation as the change of phase obtained by parallel transport around an infinitesimal closed circuit in the base space.

Thus we have managed to construct a gauge theory in which there is no concept of 'locality ' as defined via points in a base space. At this point, in view of the fact that we have obtained similar formulae for all the relevant objects appearing in the this new approach and the more traditional one,, one may ask whether the latter is more general. To answer this question (negatively, it turns out), we examine the algebraic structures involved, viz $\mathcal{U}$ and $\mathcal{B}$. Now it is a theorem well known in alebraic topology, due to Gel'fand, which tells us that (roughly speaking) any commutative algebra can be considered as an algebra of functions over some space, whose derivations are vector fields in that space: Δ is a derivation of $\mathcal{U}$ if for $a, b \in \mathcal{U}$ we have

$$\Delta(ab) = (\Delta a)b + a(\Delta b) \tag{1.55}$$

Now we saw earlier that the generators d_l of $\mathcal{B}$ act on $\Lambda \in \mathcal{U}$ by commutation [see 1.52 above], and are therefore deivations of $\mathcal{U}$. So, by Gel'fand's result, we are led to the conclusion that the d_l's are vector fields in the space over which $\mathcal{U}$ is an algebra of functions. So we see that by the commutative nature of the algebra $\mathcal{U}$, we are forced to conclude that the two approaches coincide. However, even in the case of non-commututativity (such as realised by ordinary Yang-Mills theory over $\mathbb{R}^4$), the generalisation we are following reduces to the previous case.

But there is hope! It turns out that if $\mathcal{U}$ is non-commutative, then some elements of $\mathcal{B}$ can actually be elements of $\mathcal{U}$, in which case they are termed 'inner derivations' of $\mathcal{U}$. Such a scenario would be fundamentally different from the ones we have considered hitherto. Note that we require $\mathcal{U}$ to be non-commutative, but not $\mathcal{B}$ necessarily. Conceptually, we may imagine $\mathcal{U}$ and $\mathcal{B}$, being operators acting on the Hilbert space $\mathcal{H}$, as composed of rows and columns labelled by indices indicating the set of basis vectors in $\mathcal{H}$. Then if $\mathcal{B}$ is not an inner derivation of $\mathcal{U}$, then one may see this as meaning that the matrices representing elements of $\mathcal{U}$ are all diagonal with respect to the indices on which $\mathcal{B}$ acts. One then regards elements of $\mathcal{U}$ as functions over this space of indices, which is also the 'base space' on which $\mathcal{B}$ acts. This is the situation sketched above, wherein we are reduced to the familiar structures. When $\mathcal{B}$ *does* contain inner derivations of $\mathcal{U}$, these elements of

$\mathcal{U}$ also operate on the indices themselves! It is more easily seen as follows: elements of $\mathcal{U}$ are functions on some base space (of ξ's say) with

$$f : \xi \rightarrow f(\xi) \in G \tag{1.56}$$

where G is an algebra of functions, and $f \in \mathcal{U}$; $g \in \mathcal{B}$ acts as follows:
$g : \xi \rightarrow \xi' \in \xi - space.$
So when elements of $\mathcal{B}$ are contained in $\mathcal{U}$, then one can no longer distinguish a space on which $\mathcal{B}$ acts exclusively. This is a rather remarkable occurrence! If carried through fully, it would mean that we can no longer distinguish the base space, labelled by ξ say, from the 'internal symmetry space', labelled by i. In other words, if we still want to think of a base space on which $\mathcal{B}$ acts, we must allow the 'internal symmetry operators', such as Λ and A_i to act on that space also. One should not be surprised to learn that the highly acclaimed string theory may be formulated in the language we have been using. There one is dealing with a one-dimensional object. However, we shall not enter into any details concerning this topic here. Suffice it to say that it does seem rather remarkable that one may tie together so many seemingly diverse topics within the framework of gauge. However, our enthusiasm may be tempered by the following consideration: when we postulated the existence of the Hilbert space $\mathcal{H}$, we should have been aware that this space is normally constructed as a space of functions over some base space, as we have outlined earlier. This is perhaps one fault in our analysis, *viz* that we have left the construction of $\mathcal{H}$ untouched by further considerations; it is worthy of further study.

1.5 Symmetry considerations

We have seen that Berry's original idea of a system acquiring a geometric phase when subjected to a cyclic, adiabatic motion may be extended as in the cases shown in 1.3. Another assumption made by Berry in his calculation of the holonomy $\alpha(C)$, was the use of an orthonormal set of basis vectors, $|n(X(T))\rangle$, which are instantaneous eigenstates of the system Hamiltonian $H(X(T))$ (see 1.5 above). One may then identify the overlap functions

$\langle n(X(t))|d|n(X(t))\rangle$ as the components of a $U(1)$ connection. By using the theory of invariant connections [47], one may override this assumption under certain conditions.

Further considerations of symmetry relate to the manner in which the presence of the Berry connection alters the scenario involving constants of motion. This involves the application of the Born- Oppenheimer approximation. This is used in the analysis of systems which admit a division into 'fast' and 'slow' sets of variables. The procedure is to first deal with the motion of the fast variables, keeping the slow ones fixed. The analysis is completed by then considering changes in the slow, pseudo-'fixed' variables. Thus we examine a (complete system) Hamiltonian

$$H = \frac{\vec{P}^2}{2M} + \frac{\vec{p}^2}{2m} + V(\vec{R}, \vec{r}) \tag{1.57}$$

with fast variables$(\vec{p}, \vec{r})$ and slow variables$(\vec{P}, \vec{R})$.

The sub-Hamiltonian for the fast variables is

$$h(\vec{R}) = \frac{\vec{p}^2}{2m} + V(\vec{R}, \vec{r}) \tag{1.58}$$

note that h is dependent on the slow variables, $\vec{R}$. We have, as before, the eigenstates, $|n; \vec{R}\rangle$, of the sub-system given by $h(\vec{R})|n; \vec{R}\rangle = \varepsilon(\vec{R})|n; \vec{R}\rangle$. The Berry connection $\vec{A}(\vec{R}) = \langle n; \vec{R}|i\nabla_{\vec{R}}|n; \vec{R}\rangle$ arises in the effective Hamiltonian for the slow variables(in the Born-Oppenheimer approximation):

$$H_{eff} = \frac{1}{2M}(\vec{P} - \vec{A}(\vec{R}))^2 + \varepsilon_n(\vec{R}). \tag{1.59}$$

This is an extremely surprising result! We see that the fast system induces into the slow system a potential energy $\varepsilon_{\vec{R}}$ and a velocity dependent interaction involving the $\vec{A}(\vec{R})$. (Perhaps it is not so surprising if one is familiar with the ideas of 'minimal coupling'.)

If the full Hamiltonian has a symmetry, then there are constants of motion commuting with it. In the effective system, this symmetry should carry through (with possible modifications). In the case of rotational symmetry, we have the associated angular momentum constant of motion $\vec{J}$. Under

rotations, we have that the slow variables transform as $R^i \rightarrow \Lambda^{ij} R_j$ with Λ^{ij} a special orthogonal matrix. For the effective potential, $\varepsilon(\vec{R})$, to be invariant, we require that

$$\varepsilon(\vec{\Lambda}\vec{R}) = \varepsilon(\vec{R}). \tag{1.60}$$

$\varepsilon(\vec{R})$ is a scalar; for the Berry connection, the simplest assumption would be to have

$$\vec{A}(\vec{\Lambda}\vec{R}) = \vec{\Lambda}\vec{A}(\vec{R}) \tag{1.61}$$

or equivalently,

$$\vec{\Lambda}\vec{A}(\vec{\Lambda}^{-1}\vec{R}) = \vec{A}(\vec{R}). \tag{1.62}$$

However, this is not so interesting: in this case the curvature vanishes. We could appeal to the 'physical' fact that the connection has certain 'undesirable' components to make things more interesting. This means that the above law, eqn. 1.61, may be changed to obtain

$$\vec{\Lambda}\vec{A}(\vec{\Lambda}^{-1}\vec{R}) = g^{-1}(\vec{R})\vec{A}(\vec{R})g(\vec{R}) - ig^{-1}(\vec{R})\nabla_{\vec{R}}g(\vec{R}). \tag{1.63}$$

Thus we require that a gauge transformed connection appear as the result of a rotation. When $\Lambda^{ij} = \delta^{ij} - \varepsilon^{ijk}n^k$, then $\delta\vec{R} = \vec{n} \times \vec{R}$, so that

$$\vec{n} \times \vec{A} - (\vec{n} \times \vec{R} \cdot \vec{\nabla})\vec{A} = \vec{\nabla}\Theta - i[\vec{A}, \Theta], \tag{1.64}$$

where Θ is given by $g = I + i\Theta$. The curvature, $\vec{B}$ comes into play as

$$(\vec{n} \times \vec{R}) \times \vec{B} = \nabla W - i[\vec{A}, W] \tag{1.65}$$

with $W = \Theta + \vec{n} \times \vec{R} \cdot \vec{A}$.

Thus the effective Born- Oppenheimer Hamiltonian H_{eff} is rotationally invariant provided eqns. (1.60) and (1.63) are satisfied. This analysis will be very useful for us later on, when we examine its application to a gravity-matter system (see section 3.5).

1.6 Hilbert Bundles

It would be somewhat amiss if we did not take this opportunity to show an explicit application of the fibre bundle concept in quantum mechanics.

To this end, it is instructive at this point to examine in greater detail the extent to which one may view quantal evolution as parallel transport and the effect this has on symmetries and the Berry phase. This leads to a reinterpretation of long-standing results, as was shown in recent work by Graudenz ([23]). Probably the most startling aspect of this work is the fact that we need make no adjustments to the foundational structure of quantum thoery, merely translating those foundations into the language of differential geometry. We shall be considering a Hilbert bundle G over a spacetime manifold M with typical fibre isomorphic to a Hilbert space $\mathcal{H}$. This bundle is constructed as an associated bundle to a principal bundle whose structure group is the group of unitary operators, $\mathcal{U} = \{U : \mathcal{H} \to \mathcal{H} \mid UU^{\dagger} = 1\}$. Now the unitary operators act on both state vectors as well as observables (Hermitian operators) and we shall see here the implications of this.

We assume that the spacetime manifold $\mathcal{M}$ is globally hyperbolic, i.e. that it may be foliated by a set of spacelike hypersurfaces S_λ, with $\lambda \in \mathbb{R}$. A physical system will be represented by a state vector Ψ_λ defined on some hypersurface S. Observers are incorporated into the theory by associating with every observer worldline C_τ a state vector $\Psi_C(\tau)$ depending on both the curve C and the curve parameter τ. One may think of τ as the observer, B's, eigentime. The evolution of the state vector is described by a Schrödinger equation:

$$\frac{d\Psi_C(\tau)}{d\tau} = H_C(\tau)\Psi_C(\tau). \tag{1.66}$$

We may define a unitary (evolution) operator by integrating this equation to get

$$\partial U_C(\tau, \tau_0) = H_C U_C(\tau, \tau_0) \tag{1.67}$$

satisfying the initial condition $U_C(\tau, \tau_0) = 1$. With each curve $D : [a, b] \to M$ we associate the operator U_D defined by $U_D(b, a)$. We impose two further conditions on these operators:

1. For two curves E, F with the staring point of F being the end-point of E, we let $F \circ E$ be the curve obtained by first following E and then F.We then require that

$$U_{F \circ E} = U_F \circ U_E \tag{1.68}$$

2. As for the 'inverse', we require that

$$U_{E^{-1}} = U_E^{-1} \tag{1.69}$$

where E^{-1} is the curve traversed in the opposite direction.

The properties listed above allow us to interpret the U as parallel transport operators. As stated above, these operators act on a fibre bundle G over $\mathcal{M}$ with projection π_G and typical fibre $G_x(:= \pi_G^{-1}(x)) \cong \mathcal{H}$. For a curve $E : [a, b] \to M$, U_E is a map of fibres, i.e. $U : G_{D_a} \to G_{D_b}$. As usual, the parallel transport operators have an infinitesimal representation in terms of a covariant derivative on the bundle. D is expressed as $D = d - H$ in a local coordinate system Φ given by $G\mid_W \xrightarrow{\Phi} W \times \mathcal{H}$ with $G_W \stackrel{def}{=} \cup_{x \in W} G_x$; $d = \partial_\mu \cdot dx^\mu$ is the differential on W, and $H = H_\mu \cdot dx^\mu$ is an operator-valued 1-form on M. In the frame Φ, we have $H_C(\tau) = H_\mu(C_\tau)\dot{C}_\tau^\mu$, so that

$$\partial_\tau U_C(\tau, \tau_0) = H_\mu(C_\tau)\dot{C}_\tau^\mu U_C(\tau, \tau_0). \tag{1.70}$$

When we change coordinates $\chi = \Phi_2 \circ \Phi_1^{-1}$, the representative $(x, \Psi) = \Phi_1(\psi)$ of a vector ψ over x in the frame Φ_1, is mapped into the representative $(x, V(x)\Psi) = \Phi_2(\psi)$ in the frame Φ_2, with V(x) a unitary operator (acting at the point x).

As for $H = H_\mu dx^\mu$, the generator of parallel translations, we have that under the change of coordinates χ,

$$H_{\Phi_2}(x) = V(x)H_{\Phi_1}(x)V(x)^{-1} + dV(x)V(x)^{-1} \tag{1.71}$$

which further satisfies

$$(d - H_{\Phi_1})(V\Psi) = V(d - H_{\Phi_2}\Psi). \tag{1.72}$$

We now examine the question of observables. We associate with each observable quantity A an (Hermitian) operator A_φ given by the mapping $\varphi : A \to A_\varphi$. The operator A_φ maps every fibre G_x into itself and so differs from the unitary opertors U_E which map fibres to fibres. The expectation value, $\langle\psi|A_\varphi|\psi\rangle/\langle\psi|\psi\rangle$, of an observable A_φ is obtained by constructing an

hermitian inner product $P(\psi, \xi) = \langle \psi, \xi \rangle$ for vectors ψ, ξ from the same fibre, G_x say[32]. We already have an inner product $[\cdot \mid \cdot]$ in the (prototype) Hilbert space $\mathcal{H}$. Then we can write for the quantity $\langle \psi | A_\varphi | \xi \rangle$ in a local frame Φ the expression

$$\langle \psi | A_\varphi | \xi \rangle = [\Phi(\psi) \mid P_\phi(x) A_\varphi \Phi(x) \mid \Phi(\xi)] \tag{1.73}$$

where P_Φ is an Hermitian operator representing the inner product P in the fibre G_x. When we transform frames $\chi = \Phi_2 \circ \Phi_1$, we get

$$P_{\Phi_2}(x) = V(x) P_{\Phi_1} V(x)^{-1} \tag{1.74}$$

$$A_{\Phi_2}(x) = V(x) A_{\Phi_1} V(x)^{-1}. \tag{1.75}$$

Now this inner product P will be assumed to be 'compatible' with the covariant derivative, which in terms of the parallel transport operators means that

$$\langle U_D \psi | U_D \xi \rangle = \langle \psi | \xi \rangle \tag{1.76}$$

for any path D and vectors ψ, ξ from the fibre over the starting point of D. (This law has a well known analogue in the language of general relativity, *viz.* the 'metric compatibility' condition on the connection that $\nabla g = 0$. Thus we may say that the parallel transport operator preserves the 'lengths' of vectors, as it should, since it is a unitary operator.)

We now proceed to consider the case of expectation values of observables for points $y \neq x$, where x is the position of the observer B; i.e. we would like to predict the values of certain quantities not in B's localé. To do so, we choose a curve $D : [a, b] \to M$ joining $x = D_a$ and $y = D_b$; this allows us to transport the observer's state vector ψ_B from x to y with the help of the parallel transport operator U_D. Then we obtain $\psi_B^D = U_D \psi_B$ for the state vector at the point y. As before, an expectation value for measurement of the observable A is given by $\langle \psi_B^D | A_\varphi | \psi_B^D \rangle$. Thus we parallel transport the state vector and allow A to act upon it (at the point y). For interesting cases, we examine the scenario where the parallel transport of ψ_B depends on the path chosen. This will, of course, evoke the notion of the curvature F (of the connection D). F is an operator- valued 2-form, taking two tangent vectors $v, w \in T_x M$ and a state vector $\psi \in G_x$ into a vector $F(v, w)\psi \in Gx$. Note

that the 'domain' and 'range' are the same in the above exactly because F is operator-valued. F tells us to what extent parallel transport round a closed loop β fails to coincide with the identity; thus

$$U_\beta = \imath d + \tau^2 F(v, w) \tag{1.77}$$

Since the *expectation value*, $\langle \psi_B^D | A_\varphi(y) | \psi_B^D \rangle$ does not depend on the path D, it would be wise to examine exactly what consequences follow from travelling along a different path. For another path E joining x and y, we thus have

$$\langle \psi_B^D | A_\varphi(y) | \psi_B^D \rangle = \langle \psi_B^E | A_\varphi(y) | \psi_B^E \rangle \tag{1.78}$$

whence

$$\langle U_D \psi_B | A_\varphi(y) | U_D \psi_B \rangle = \langle U_E \psi_B | A_\varphi(y) | U_E \psi_B \rangle \tag{1.79}$$

Since this hold for all y, $\forall \psi_B \in G_x$ and for all curves D and E joining x and y, we have

$$< \xi \mid U_\alpha A_\varphi(y) U_\alpha^{-1} - A_\varphi(y) \mid \xi >= 0 \tag{1.80}$$

where $\xi = U_D \psi_B \in G_y$ (since U_D is a map of fibres). Thus we finally obtain

$$[\mathbf{U}_\alpha, \mathbf{A}_\varphi(\mathbf{y})] = \mathbf{0}. \tag{1.81}$$

Since the set of all parallel transport operators U_α along closed loops α at y form the holonomy group $H(y)$ (of the connection D) at y, the relation above tells us that $H(y)$ has to commute with all observables at y (i.e all point observables). The U_α transform states into states in a non-trivial manner, while leaving observables invariant. This allows us to interpret the parallel transport as local symmetry transformations! Recall our discussion on gauge theories, wherein we explained how quantities are deemed to be 'observables' by virtue of the fact that they are invariant under a 'gauge transformation' (and have to commute with the gauge group). Thus it would be of considerable interest if one could relate the holonomy group, which acts here as a symmetry group, to some kind of internal symmetry space or gauge group. We shall have more to say about this later.

We continue by examining the role of curvature in the result derived above. First we define for a curve E joining x and y the operator A_φ^E on

G_x by the expression $U_E^{-1}A_\varphi(y)U_E$. Thus A_φ^E maps G_x into G_x and further satisfies the following:

$$\begin{aligned}\langle\psi_B^E|A_\varphi(y)|\psi_B^E\rangle &= \langle U_E\psi_B|A_\varphi|U_E\psi_B\rangle \\ &= \langle\psi_B|U_E^\dagger A_\varphi(y)U_E|\psi_B\rangle \\ &= \langle\psi_B|U_E^{-1}A_\varphi(y)U_E \mid \psi_B\rangle \\ &= \langle\psi_B|A_\varphi^E|\psi_B\rangle \qquad (1.82)\end{aligned}$$

Consider a closed loop β at x. Then $\alpha = E\cdot\beta\cdot E_{-1}$ is a closed loop at y. The condition $[U_\alpha, A_\varphi(y)] = 0$ may be rewritten as $[U_\beta, A_\varphi^E] = 0$. Since $U_\beta = \imath d + \tau^2 F(v,w)$ as shown earlier, we obtain

$$[U_\beta, A_\varphi^E] = [\imath d, A_\varphi^E] + \tau^2[F_{\mu\nu}(x), A_\varphi^E] = 0; \qquad (1.83)$$

but since the identity map commutes with all observables, we obtain that

$$[F_{\mu\nu}(x), A_\varphi^E] = 0, \qquad (1.84)$$

i.e. the curvature tensor at x has to commute with all observables parallel transported to x along arbitrary curves. (One may ask whether this is a condition on the curvature or on the observables.) One may define a family of curves ε_y joining y and x by $\varepsilon_y(\tau) = h(y,\tau)$, where $h(y,\tau)$ is a homotopy of the map $U \to M$. $U \subset M$ is a distant region of the spacetime M, and the curves ε admit a local representation of this region at x: define $A_{\varphi h}(y) = A_\varphi^{\varepsilon_y} = U_{\varepsilon_y}A_\varphi(y)U_{\varepsilon_y}^{-1} = A_\varphi^{\varepsilon_y^{-1}}$. $A_{\varphi h}(y)$ is an operator on the fibre G_x. The expectation value for A at $y \in U$, given that B's state vector is ψ_B, is just $< \psi_B \mid A_{\varphi h}(y) \mid \psi_B >$ In a local frame, the operators $A_{\varphi\varepsilon_{y|\tau}}$ satisfy the equation

$$\partial_\tau A_{\varphi\varepsilon_{y|\tau}} = [H_\mu(\varepsilon_\tau)\dot{\varepsilon_\tau^\mu}, A_{\varphi\varepsilon_{y|\tau}}]. \qquad (1.85)$$

This may be seen as the Heisenberg equation of motion for the operators $A_{\varphi\varepsilon_{y|\tau}}$.

Graudenz applies this formalism to a number of cases of interest; in particular, he treats the case of spacetimes with closed timelike curves. The problems associated with this situation include the possibility that the evolution may be non-unitary[16]. However, by maintaining evolution with the

use of a Schrödinger equation, he hopes to maintain explicit unitarity. One may well imagine how the formalism developed here of the effect of unitary operators, U_β on state vectors and operators may be adapted to the case of fields defined in the presence of closed timelike curves. We shall not enter into such a discussion, but refer the reader to the original literature[23]. Rather, we turn our attention to the interesting results obtained by Graudenz in the course of his investigation. As he points out, it remains an open issue the question of the physical interpretation of the symmetry result derived earlier. Tied-in with this is another interesting question concerning the relation of the holonmy obtained here to the situations invoving the presence of the Berry phase. In the latter, a Hamiltonian $H(\lambda(t))$ describes the temporal evolution of the system, while for the present formalism "...the objects describing the time evolution are the generators of the parallel transport operators themselves", i.e. $U \rightarrow D = d - H$. In addition, the Berry phase was originally formulated over parameter space, while the phase here occurs over spacetime (we shall have more to say about this in section 3.5). Now it is precisely because the phase is now not observable that one has the luxury of interpreting the result above to imply that the holonomy group as symmetry group. We point out another possibility as to how the phase differs from the Berry phase case (and thus possibly also why the latter is observable): Simon's original interpretation of the Berry phase involved the latter as an holonomy associated with a complex line bundle over parameter space; this bundle inherited a connection from the space it was embedded in, viz $\mathcal{M} \times \mathcal{H}$. In Graudenz's work, we are dealing with a *full* Hilbert space as fibre, and not the line bundle. Thus the $U(1)$ symmetry present in the Berry phase case (which, in the original derivation, comes from identifying the eigenstates of the Hamiltonian) is no longer explicit, and perhaps not even present.

1.7 Gravitational Phase

Notwithstanding the differences between gravity and Yang-Mills type gauge fields mentioned earlier, it is interesting to examine how the former may influence quantum phase considerations. We shall exhibit here some of the

conceptual issues involved[4].

Recall that for an arbitrary Yang-Mills gauge field, the phase factor,

$$F_\gamma = Pe^{-\frac{ig}{\hbar c}\oint_\gamma A_\mu^k T_k dx^\mu} \tag{1.86}$$

determines the phase shift in quantum mechanical wavefunctions; P denotes path ordering, γ is a closed curve, T_k is a generator of the Lie algebra and A_μ^k is the Yang-Mills gauge potential. F_γ is an element of the Lie group (arising, heuristically, from the exponential map). For a surface element $d\sigma^{\mu\nu}$ enclosed by γ, we may evaluate F_γ to yield

$$F_\gamma = 1 = \frac{ig}{2\hbar c} F_{\mu\nu}^k T_k d\sigma^{\mu\nu} \tag{1.87}$$

where $F^k = dA^k - gC_{ij}^k A^i \wedge A^j$ is the field strength. Now the phase shift of a particle due to the gravitational field is determined by

$$F_\gamma = Pe^{\frac{i}{\hbar}\int_\gamma (e_\mu^a P_a + \frac{1}{2}\Gamma_\mu^{ab} M_{ab}) dx^\mu} \tag{1.88}$$

Since, as mentioned above, the phase factor is an element of the gauge group, F_γ is here an element of the Poincaré group; now, however, the path γ need no longer be a closed path in spacetime. This is notably in contrast to the Yang-Mills scenario shown above, and results from the difference mentioned in section 1.4.1 between General Relativity and Yang-Mills theories: the translational gauge symmetry of the former is said to be broken by the existence of certain Higgs fields. We shall identify these fields in the following. In the above, P_a and $M_{ab}, a, b = 0, 1, 2, 3$ are the energy-momentum and angular momentum operators which generate the representation of the Poincaré group under consideration, and Γ_μ^{ab} are the connection coefficients associated to the frame field e_a^μ (section of the frame bundle) used by local observers. The field e_μ^a is dual to the e_a^μ; if the latter are orthonormal,then

$$e_a^\mu e_b^\nu g_{\mu\nu} = \eta_{ab} \tag{1.89}$$

For a spinor, the values of the wavefunction are observed by an observer using the frame e_a^μ. Then equation above indicates that in addition to a pahse from

the energy-momentum part (viz $e^a_\mu P_a$), the spinor field is parallel transported (using $\frac{1}{2}\Gamma^{ab}_\mu M_{ab}$). For a spinless particle, $M_{ab} = 0$ and so the gravitational phase obtained is thus

$$\varphi = \frac{1}{\hbar c}\int_\gamma e^a_\gamma p_a \tag{1.90}$$

where p_a are the eigenvalues of the energy momentum operator P_a. As mentioned earlier this phase is observeable even for an open curve, which comes about due to the presence of the field e^a_μ, as we shall demonstrate: **(A)** when we compare the phase factors for Yang-Mills and general relativity as shown above, we note immediately that the terms $A^k_\mu T_k$ and $e^a_\mu P_a$ are analogous; thus the e^a_μ may be identified as a connection/gauge potential associated with the translation group (the P_a, being energy-momentum operators, acting as generators of that group, as noted above). Then, e^a_μ together with Γ^{ab}_μ may be regarded as giving us the connection in the affine bundle. The curvature of this connection is obtained by evaluating F_γ in eqn. 1.88 above for an infinitesimal closed curve γ:

$$F_\gamma = 1 + \frac{i}{2\hbar}(Q^a_{\mu\nu}P_a + \frac{1}{2}R^{ab}_{\mu\nu}M_{ab})d\sigma^{\mu\nu} \tag{1.91}$$

where $Q^a = de^a + \Gamma^a_b \wedge e^b$ is the torsion, and $R^{ab} = \Gamma^{ac} \wedge \Gamma_{cb}$ is the curvature. In this way, gravity may be seen to be a gauge field associated to the Poincaré group;

(B) e^a_μ is the pullback of the solder form with respect to the local section e^μ_a in the bundle of frames [44]. Then the ($\mathcal{G}$-valued) 1- form $e^a_\mu P_a$ acts on the tangent vector to γ to give an element in the Lie-algebra of the translation group; the latter is an observable acting on the Hilbert space, which has as (approximate) eigenvalue the rate of change of phase along γ. The *total* phase change is obtained by integrating over γ, which results in the gravitational phase quoted above.

(C) e^a_μ is like the 'square root' of the metric, since $e^a_\mu e^b_\nu \eta_{ab} = g_{\mu\nu}$.

With regard to the (A), (B) and (C) above, we may make the following comments:

- If we see e^a_μ as a connection associated with the translation group, then there is no restriction on its value at any spacetime point; in fact, it can

be made to vanish along any differentiable curve by a suitable choice of gauge. Then the 'gravitational field' has the full gauge symmetry of the affine group $A(4, \mathbb{R})$. By virtue of the fact that the holonomy group is a subgroup of the Poincaré group, this means that only generators of the Poincaré Lie algebra may occur in F_γ in eqn. 1.91 above. This means that in order to be observable, the 'gravitational phase' would have to be measured for a closed curve γ.

- By viewing e^a_μ as the pullback of the solder form, the matrix e^a_μ is constrained to be non-singular. The gauge symmetry group becomes $GL(4, \mathbb{R}) \subset A(4, \mathbb{R})$, with Γ^{ab}_μ acting as the connection. It is the breaking of the translational gauge symmetry that enables the phase in eqn 1.91to be observed. Now however, we are not left with a gauge theory with $GL(4, \mathbb{R})$ as the internal symmetry space; rather the soldering nature of the e^a_μ provides us with richer structure (essentially due to the possible invariants which may be constructed with the soldering form in place, as opposed to Yang-Mills gauge theories, where no such soldering exists). 'Local gauge transformations' now turns out to be nothing other than spacetime dependent transformations of the frame field e^μ_a. Since the solder pullback e^a_μ is dual to the frame field, it means that e^b_μ transforms as a tensor; thus the phase is invariant under the above gauge transformations and so is observable.

- When we introduce the metric via the soldering form as in (C) above, this causes a further breaking of symmetry down to the Lorentz group $O(3, 1, \mathbb{R}) \subset GL(4, \mathbb{R})$, now leaving the metric invariant. Thus both the soldering form, e, and the metric, g, may be seen as Higgs fields.

1.8 Quo Vadis BP?

The Berry phase was originally motivated out of considerations in molecular physics. But its importance and application has since stretched beyond that. As we have seen so far, Berry's original derivation of the phase arising from adiabatic motion in a close loop in parameter space has been extended to

both non-cyclic as well as non-adiabatic motion, as shown in the results of [1] and [37]. Furthermore, symmetry considerations show that it is possible to calculate a Berry phase even without the use of the eigenstates of the Hamiltonian. We then elaborated on several areas where the phase is of importance. We shall collect here a few remarks concerning the work we have exhibited so far, as well as suggesting interesting avenues where further investigation may reveal unexplored approaches/views to the phase which may be of value.

Our concern is with a somewhat unorthodox view of the Berry phase: we have seen that the Berry/geometric phase is the integral of a Lie-algebraic valued one-form. As for the dynamical phase, in a differential geometric setting, the Hamiltonian may be seen as a Lie-algebraic valued zero-form or function. Thus the dynamical phase $\gamma_d = \int \langle \psi \mid d\psi \rangle$ is the integral of a zero-form, while the Berry phase $\gamma_b = \int A_\mu dx^\mu$ is the integral of a one-form. With this view, and the arrogance of hindsight, one is led to the conclusion that the Berry phase is an 'obvious' one in the sense of being the next step after the dynamical phase. Then one is led to consider the next 'obvious' step as the consideration of an integral over a two-form, $X_{\mu\nu}$, say. However, we have postulated this only from a mathematical point of view, and do not claim that there is, as yet, any physically plausible role for this two-form. But, on the other hand, it is not so easy to dismiss these claims out of hand. Perhaps the major reason why the Berry phase remained undiscovered for more than fifty years after the discovery of its (original) starting point, *viz* the adiabatic theorem, is due to the fact that quantum mechanical phases were generally considered to possess no physically measureable properties, and could thus be ignored. Thus, one becomes more wary of the dismissal of certain effects on the basis that they are 'unverifiable' or 'just a phase'.

Now one may seek to continue this procedure and postulate the existence of (the integral of) a three-form or even a four-form. One may even adopt the view that this process could continue indefinitely, each new step resulting in another 'correction to the phase'. (Note, however, that our remarks here differ drastically from the proposals for iterative solutions for the phase and consequent renormalisation obtained by Berry in [40]. Also, the

two-form we are proposing here differs, at least conceptually, from the two-form which arises in the analysis of systems using the Born- Oppenheimer approximation[28]: there a two-form originates from the effect of the light system on the heavy on. What we are proposing here comes from the influence of the latter on the former, as in the original derivation of the 1-form by Berry[9].)

Indeed, one may postulate this as a Taylor series of sorts for the Hamiltonian, H, wherein one sees the Berry connection, A, as the first term, arising from the application of the exterior derivative as follows: $A = dH$, or in component form:

$$A_\mu = \frac{\partial H}{\partial x^\mu} dx^\mu \tag{1.92}$$

In fact, this view of the Berry connection is adopted by a number of authors (see especially [28]), viz that one is merely calculating 'higher order corrections' to the dynamics as determined by the Hamiltonian. However, we take exception to such a view on the following basis:

1. Even though we may write $A = dH$ locally, it must be kept in mind that it is precisely a *local* result. When written in the form $A = dH$, we are in fact, postulating that as a one-form, the Berry connection A is exact (i.e. it can be expressed as the differential of a zero form, H). The conditions under which such a situation holds globally may be examined from the point of view of the following result [32]:

   Lemma 1.3 *(Poincaré): In $\mathbb{R}^n$ a form is exact $\Longleftrightarrow$ it is closed.*

   We know that this result woud not be expected to hold for a general manifold, but the relation $A = dH$ could always be solved locally. The implications of this result holding globally may be seen from the application of the exterior derivative operator to the above relation: $dA = d^2H = 0$ (by property of d) so that A is, in fact, 'constant', i.e. has zero (exterior) gradient. For this result to hold globally would mean that we have a manifold of zero curvature (since curvature $F = dA$).

Thus we are led to the conclusion that the view of the phase arising from the connection A as a 'correction' to the dynamical phase (seen as arising from the Hamiltonian H) is a slightly misleading one. Besides, the phase is itself a global object, being the integral of A over a region enclosed in the curve C in parameter space, $\gamma_b = \int_C \langle n \mid dn \rangle$. Questions concerning the 'disappearance' of γ_b are dealt with in [22].

2. Thus one should view the existence of the connection and associated phase as essentially independent of the Hamiltonian, but also part of a larger scheme. This scheme would include the contributions arising from considerations of the two-form referred to above. There we mentioned that this object, among others, may be seen as forming part of a 'Taylor series' for the Hamiltonian. But besides the fact that this series would truncate after the first term (due to the fact that the first term arises as the application of the d operator to the Hamiltonian H, while the second term would have to be zero, being the subject of the application of d^2 to H), we also have that a 'natural cut-off' arises in the following sense: it is all very well for the purported objects to exist, but their effects are measured via integrals; now when discussing forms over 4-dim spaces, for example, the integrals concerned are seen to vanish for forms of order greater than 4. This is an elementary result in differential geometry.

The issue of these forms may be an interesting question to pursue then in possible future work. With this we take leave of the analysis of geometric phases and quantum mechanics, and turn our attention to the case of fields, ultimately returning to the extension of Berry's results in section 3.5.

Chapter 2

Quantum Field Theory

2.1 Flat space quantization

Quantum field theory has impeccable credentials as far as experimental verification is concerned, but for our purposes (which is mainly its application in curved spacetime), the theory is sometimes unsound. So, although we shall not be utilising the standard approach to field theory for our main results, we briefly review that approach here chiefly to highlight its deficiencies. We will shortly examine the revisions of the theory we require.

We shall restrict attention to a Klein-Gordon scalar field, φ in Minkowski spacetime $(\mathbb{R}^4, \eta_{ab})$.

The action of the Klein-Gordon field in $(\mathbb{R}^4, \eta_{ab})$ is

$$S = -\frac{1}{2}\int(\partial_\mu\varphi\partial^\mu\varphi + m^2\varphi^2)d^4x \tag{2.1}$$

The Lagrangian field density for the field is

$$\mathcal{L} = [\frac{1}{2}\int\dot{\varphi}^2 - \vec{\nabla}\varphi^2 - m^2\varphi^2]d^3x, \tag{2.2}$$

where the dot indicates the time derivative. The field, φ is evaluated on a t = constant hypersurface,and is treated as the configuration variable. The conjugate momentum is $\pi = \frac{\partial L}{\partial\dot{\varphi}} = \dot{\varphi}$.

At this stage, most textbook treatments follow the cumbersome route of "putting the system in a box", i.e., replacing infinite Euclidean space, $\mathbb{R}^3$, by a flat 3-torus, T^3, of side L. Besides the fact that this procedure is not Lorentz covariant [46], it also hides the real mathematical nature of the fields. But for now, by placing the system in the aforementioned box, we may expand φ in a Fourier series:

$$\varphi(t,\vec{x}) = \frac{1}{\sqrt{L^3}}\Sigma_{\vec{k}}\varphi_{\vec{k}}(t)e^{i\vec{k}\vec{x}}, \tag{2.3}$$

where the wave vector, $\vec{k}$, has the form $\vec{k} = \frac{2\pi}{L}(n_1, n_2, n_3)$. (If we had insisted on using $\mathbb{R}^3$ instead, then we would have worked with integrals of the form $(\frac{L}{2\pi})^2 \int d^3x$ instead of the sum $\Sigma_{\vec{k}}$.) The coefficients $\varphi_{\vec{k}}(t)$ are given by

$$\varphi_{\vec{k}}(t) = \frac{1}{\sqrt{L^3}}\int \varphi(t,\vec{x})e^{-i\vec{k}\vec{x}}d^3x, \tag{2.4}$$

and satisfy $\overline{\varphi_{\vec{k}}} = \varphi_{-\vec{k}}$ (since $\varphi(t,\vec{x})$ is real). The Lagrangian in equation 2.2 is rewritten as

$$\mathcal{L} = \Sigma_{\vec{k}}(\frac{1}{2}|\dot{\varphi}_{\vec{k}}|^2 - \frac{1}{2}\omega_{\vec{k}}^2|\varphi_{\vec{k}}|^2, \tag{2.5}$$

with $\omega_{\vec{k}}^2 = \vec{k}^2 + m^2$.

Our familiarity with the harmonic oscillator of elementary quantum mechanics (see e.g. [38]) leads to the conclusion that the Klein-Gordon field may be formally viewed as an infinite collection of decoupled, time-independent, harmonic oscillators. We may construct a Hilbert space for a system of n such oscillators by taking the tensor product of the Hilbert spaces for the individual oscillators. However, for the case we are considering of a countably infinite system of such oscillators, the (infinite) tensor product suffers from several drawbacks: chief among these are the fact that it is not separable, and the representation it provides of the canonical commutation relations is reducible. Thus we seek an alternative Hilbert space, and choose for the infinite system a Fock space (see section 2.1.2) which is based on a Hilbert space, $\mathcal{H}$, of complex classical solutions to the Klein-Gordon equation of motion: $\partial^\mu\partial_\mu\varphi - m^2\varphi = 0$. An orthonormal basis for $\mathcal{H}$ is provided by "plane

wave" solutions of the form

$$\psi_{\vec{k}} = \frac{1}{\sqrt{2\omega_{\vec{k}} L^3}} e^{i\vec{k}\vec{x} - i\omega_{\vec{k}} t} \tag{2.6}$$

Now in identifying the field as an infinite collection of oscillators, we in effect, place the role of "position observables" onto the quantities $\sqrt{2}\mathrm{Re}(\varphi_{\vec{k}})$ and $\sqrt{2}\mathrm{Im}(\varphi_{\vec{k}})$ [since by comparing the Lagrangians for the single and infinite case, $\mathcal{L}_1 = \frac{1}{2}\dot{q}^2 - \frac{1}{2}\omega^2 q^2$ and $\mathcal{L}_\infty = \Sigma_{\vec{k}}[\frac{1}{2}\dot{\varphi}_{\vec{k}}\dot{\bar{\varphi}}_{\vec{k}} - \frac{1}{2}\omega_{\vec{k}}^2 \varphi_{\vec{k}}\bar{\varphi}_{\vec{k}}]$ respectively, we see heuristically that q 'splits in two', into $\varphi_{\vec{k}}$ and $\bar{\varphi}_{\vec{k}}$; in addition, the factors of $\sqrt{2}$ in the above quantities arise from the fact that $\bar{\varphi}_{\vec{k}} = \varphi_{-\vec{k}}$.]

So, like the case of a single oscillator, where we ascribed an annihilation operator, a, to each position variable, q, so here we ascribe annihilation operators, $b_{\vec{k}}$ and $c_{\vec{k}}$ to each of the position operators for the field case, viz. $\sqrt{2}\ \mathrm{Re}(\varphi_{\vec{k}})$ and $\sqrt{2}\ \mathrm{Im}(\varphi_{\vec{k}})$, respectively. Then we can write for the position observables in the Schrödinger picture:

$$\sqrt{2}\ \mathrm{Re}(\varphi_{\vec{k}}) = \frac{1}{\sqrt{2\omega_{\vec{k}}}}(b_{\vec{k}} + b_{\vec{k}}^\dagger) \tag{2.7}$$

$$\sqrt{2}\ \mathrm{Im}(\varphi_{\vec{k}}) = \frac{1}{\sqrt{2\omega_{\vec{k}}}}(c_{\vec{k}} + c_{\vec{k}}^\dagger) \tag{2.8}$$

We then define the operator $a_{\vec{k}}$ by $a_{\vec{k}} = \frac{1}{\sqrt{2}}(b_{\vec{k}} + ic_{\vec{k}})$ from which follow the commutation relations: $[a_{\vec{k}}, a'_{\vec{k}}] = 0$ and $[a_{\vec{k}}, a_{\vec{k}}^\dagger] = \delta_{\vec{k}\vec{k}'} I$.

Up till now, the object $\varphi_{\vec{k}}$ has been a (complex) classical observable; with the definitions above, we may establish a corresponding quantum observable $\hat{\varphi}_{\vec{k}}$ as

$$\hat{\varphi}_{\vec{k}} = \frac{1}{\sqrt{2\omega_{\vec{k}}}}(a_{\vec{k}} + a_{-\vec{k}}^\dagger). \tag{2.9}$$

Then we obtain the following expression for the operator $\hat{\varphi}_{\vec{k}}(\vec{x})$ representing the value of the field φ at some spatial point $\vec{x}$:

$$\hat{\varphi}(\vec{x}) = \frac{1}{\sqrt{L^3}}\Sigma_{\vec{k}} e^{i\vec{k}\vec{x}}\hat{\varphi}_{\vec{k}} \quad = \quad \Sigma_{\vec{k}}\{\frac{e^{i\vec{k}\vec{x}}}{\sqrt{2\omega_{\vec{k}} L^3}} a_{\vec{k}} + h.c.\},$$

whre $h.c.$ indicates Hermitian conjugate. The Heisenberg picture operator, $\hat{\varphi}(t,\vec{x})$, is given by

$$\hat{\varphi}(t,\vec{x}) = \frac{1}{\sqrt{2\omega_{\vec{k}}L^3}}\Sigma_{\vec{k}}\{e^{i\vec{k}\vec{x}-i\omega_{\vec{k}}t}a_{\vec{k}} + h.c.\}, \tag{2.10}$$

which is of the form $\hat{\varphi}_{\vec{k}}(t,\vec{x}) = \Sigma\{\psi_i(t,\vec{x})a_i + \hat{\psi}_i(t,\vec{x})a_i^\dagger\}$, where ψ is an element of the orthonormal basis of $\mathcal{H}$ and a_i is the associated annihilation operator. This association is 'natural', emulating the single harmonic oscillator case, where $q_H(t) = \sqrt{\frac{1}{2\omega}}(e^{-i\omega t}a + e^{i\omega t}a^\dagger)$. Although it appears that we have arrived at 2.10 quite easily, it must be pointed out that we did so at the expense of mathematical rigour. This is not just a formal concern: the infinite series defining $\varphi(t,\vec{x})$ does not converge; thus we do not obtain an observable corresponding to the value of the field at the spacetime point $(t,\vec{x})$. So even though we passed over to the 'box limit', this hasn't helped too much: we replace continuous integrals over plane wave modes with a discrete sum over a basis of normalised states (ψ_i), to no avail. Further difficulties arise from considerations of energy-momentum. These concern the so-called vacuum energy divergence. The vacuum, $|0>\in \mathcal{F}(\mathcal{H})$, is delineated by the condition $a_{\vec{k}}|0>=0 \quad \forall\vec{k}$, just as in the case of a single oscillator. In general, arbitrary elements of $\mathcal{F}(\mathcal{H})$ are written as

$$\Lambda = (\lambda_0, \lambda_1, \lambda_2, ...), \tag{2.11}$$

where $\lambda_0 \in \mathbb{C}$ is interpreted as the amplitude to be in the vacuum state, $\lambda_1 \in \mathcal{H}$ is the one-particle amplitude, and so on. Thus $|0>$ is defined as the vector with $\lambda_0 = 1$, and $\lambda_i = 0$, $\forall i \geq 1$. The many particle states are constructed by repeated application of the creation operator, $a_{\vec{k}}^\dagger$ to $|0>$. The action of these operators is defined as follows: for $\phi \in \mathcal{H}$,

$$\hat{a}^\dagger(\phi)\Lambda = (0, \lambda_0\phi, \sqrt{2}\lambda_1 \otimes_s \phi, \ldots). \tag{2.12}$$

So, for example, the one-particle state, normally denoted $|1>$, is obtained as $|1> = \hat{a}^\dagger(\phi)|\lambda_0, 0, 0, ...> = |0, \lambda_0\phi, 0, 0, ...>$. (The rather interesting question of the extent to which this field state, $|1>$ acts as a wavefunction

for the 'single particle', is examined in [45].) As for the annihilation operator, $\hat{a} : \mathcal{F}(\mathcal{H}) \rightarrow \mathcal{F}(\mathcal{H})$, its action on an arbitrary vector, $\mathcal{L} \in \mathcal{F}(\mathcal{H})$, is defined as:

$$\hat{a}(\phi)\Lambda = (\bar{\phi}\lambda_1, \sqrt{2}\bar{\phi}\lambda_2, \ldots). \tag{2.13}$$

Since we have interpreted the vacuum state, $|0>$ as the zero-particle state, we would expect it to have zero energy-momentum. This assumption is false, as borne out taking the appropriate expectation values of the energy-momentum operator. This operator is the quantum version of the classical energy-momentum tensor, $T_{\mu\nu}$. For the field, φ in $(R^4, \eta_{\mu\nu})$, we have

$$T_{\mu\nu} = \varphi_{,\mu}\varphi_{,\nu} - \frac{1}{2}\eta_{\mu\nu}\eta^{\lambda\delta}\varphi_{,\lambda}\varphi_{,\delta} + \frac{1}{2}m^2\varphi^2\eta_{\mu\nu}. \tag{2.14}$$

We obtain the Hamiltonian operator, H, as $H = \int_t T_{tt} d^3x$, with quantum analogue,

$$\hat{H} = \frac{1}{2}\Sigma_{\vec{k}}(a^\dagger_{\vec{k}}a_{\vec{k}} + a_{\vec{k}}a^\dagger_{\vec{k}})\omega \tag{2.15}$$

Similarly, for the momentum operator, $P_i = \int_t T_{ti} d^3x$, $\hat{P} = \Sigma_{\vec{k}} a^\dagger_{\vec{k}} a_{\vec{k}}$. As far as the latter is concerned, the vacuum expectation value is zero: $< 0|\hat{P}_i|0 >= 0$, $i = 1, 2, 3$.
However, by rewriting the Hamiltonian as $\hat{H} = \Sigma_{\vec{k}}(a^\dagger_{\vec{k}}a_{\vec{k}} + \frac{1}{2})\omega$, we obtain

$$< 0|\hat{H}|0 >=< 0|0 > \Sigma_{\vec{k}}\frac{1}{2}\omega = \Sigma_{\vec{k}}\frac{1}{2}\omega. \tag{2.16}$$

This quantity is actually infinite! To see this explicitly, we first convert from 'box' to integral form by replacing $\Sigma_{\vec{k}}$ by $(\frac{L}{2\pi})^{n-1}\int d^3x$, so that

$$\begin{aligned} < 0|\hat{H}|0 > &= \frac{1}{2}(\frac{L}{2\pi})^{n-1}\int \omega d^3k \\ &= (\frac{L}{2\pi})^{\frac{n-1}{2}}\frac{1}{\Gamma(\frac{n-1}{2})}\int_0^\infty \sqrt{k^2 + m^2}k^2 dk \end{aligned} \tag{2.17}$$

which diverges (due to the poles in the Γ function, as well the 'ultraviolet' end of the integral). To bypass this difficulty, one releases the ruse of renormalisation. Physically, we say that energy is not observable (at

least in $(\mathbb{R}^4, \eta_{ab})$), so that we are free to redefine the zero-point. This is realised by introducing an operation called normal ordering (on operators), denoted by ": *operator* :", whereby all annihilation operators are shifted to the right of creation operators where they occur together in products. Thus $: \hat{H} := \Sigma_{\vec{k}} a^{\dagger}_{\vec{k}} a_{\vec{k}} \omega$, so that $< 0| : \hat{H} : |0 >= 0$.

We are led to conclude that even though the vacuum vector contains no particles, it nonetheless seems to possess certain unexpected properties (which we have somewhat ruthlessly swept under the rug of renormalisation); this should alarm one to the possibility that particles are not all there is to field theory. As Unruh as warned: "At all times one must ... remember that the fundamental theory is not a particle theory, and that describing the system in terms of particles may be misleading " [45]. However, in certain situations, the concept of particle may be useful. We shall examine this in the following subsection. Before doing so, however, we pause here to take stock: when discussing field theory, we immediately looked for (and found) an analogy with the case of a single harmonic oscillator. We thus deduced that field theory was nothing other than a theory of many particles. We glossed over some of the more 'technical' questions which arose from:
(I): issues of convergence, preferring instead to ponder the nature of observables.
We then found to our surprise **(II)**: that there were some difficulties associated with the seemingly innocuous 'vacuum', $|0 >$. We shall deal with the issues (I) and (II) (in reverse order) in the following. They are not entirely unrelated: if we manage to put the field on a sound footing without recourse to a particle interpretation, then we should be able to solve (I), and possibly then also make (II) void.

2.1.1 Particles

To define the notion of particles, we require that physical laws are invariant under Poincaré transformations, i.e. Lorentz transformations and spacetime translations. In terms of what we have done thus far, this means that fields and states are to be characterised by their transformation properties under the Poincaré group. Specifically, we use the (quantum) operator versions of

P_μ and $J_{\mu\nu}$ to generate unitary representations of this group, where $P_\mu = \int_t T_{ti} d^3x$ as before, and $J_{\mu\nu}$ is the angular momentum tensor. $\hat{P}^\mu$ and $\hat{J}_{\mu\nu}$ are hermitian operators acting on the space of states. Different irreducible representations of the Poincaré group are labelled by the group's Casimir operators, which are $P^2 = P^\mu P_\mu$ and $W^2 = W^\mu W_\mu$. W_μ is the Pauli-Lubánski vector, given by $W_\mu = -\frac{1}{2}\varepsilon_{\mu\nu\lambda\sigma} J^{\nu\lambda} P^\sigma$. [In general, $W_\mu P^\mu = 0$]. We are not especially motivated to study the explicit construction of the representations, except to say that it is accomplished by Wigner's little group method, wherein we consider an eigenstate $|p>$ of P^2, W^2 and P^μ obeying $P^\mu|p>= p^\mu|p>$. The idea is to fix $p^2 = p^\mu p_\mu$ and then to generate basis states for all values of p^μ in terms of those for a single fixed vector.

A particle is then defined as a state of a quantum field that transforms under elements of the Poincaré according to a definite irreducible representation. This implies (from the labelling of such representations in terms of P^2 and W^2) that a particle has a definite mass and spin. So, as we have have stressed, it should be obvious that this framework relies heavily on the Poincaré group, which is the symmetry group of Minkowski space. In general (curved) spacetimes then, one would expect the particle framework to suffer severely. (We have avoided questions relating to the C, P and T transformations; theses are dealt with in [52].) This is due to the fact that in these spacetimes, there is normally no simple symmetry group which characterises the spacetime geometry. In particular, the mode decomposition of the field, ϕ, in terms of the plane wave modes, $e^{i\vec{k}\vec{r}-i\omega t}$, (which are naturally associated with the Poincaré group, being eigenfunctions of the $\frac{\partial}{\partial t}$ operator, since $\partial_t e^{i\vec{k}\vec{r}-i\omega t} = -i\omega e^{i\vec{k}\vec{r}-i\omega t}$), is no longer viable, since in a curved spacetime, such simple exponentials are absent.

Now it may be thought that in foregoing the luxury of the Poincaré group, we also forfeit the very concept of Hilbert space. After all, the latter is usually conceived of as a representation space for the Poincaré group. Intuitively, one would expect that in losing the simple exponential modes, $e^{i\vec{k}\vec{r}-i\omega t}$, we not only lose a basis for $\mathcal{H}$, but also disrupt the conceptual (and constructive) nature of $\mathcal{H}$. All these fears are allayed by the realisation that the Hilbert space is, in effect, a local construction. We saw in section 1.1

that one constructs $\mathcal{H}$ from a space of sections. There we said that in the case of a twisted bundle (which we shall intuit as associated with a curved spacetime), we may use a space of local sections. As pointed out, this leads us to the interpretation which finds its realisation in the notions expressed above, viz. that a simple particle interpretation is no longer possible in a non-trivial spacetime.

Thus it is no great loss to abandon the exponential modes; they are globally defined and furthermore, are only a (sometimes) convenient basis for $\mathcal{H}$. We can think of the situation heuristically as follows: in differential geometry, the construction of a tangent space follows naturally from the manifold structure, and in particular, does not require additional structure such as the metric; $\mathcal{H}$ is to be thought of in a similar manner. Then just as we use coordinates as a convenient label for vectors, so we use the exponential modes to define a basis for $\mathcal{H}$.

When we deal with curved spacetime, the lack of a 'natural' basis, such as the $e^{i\vec{k}\vec{r}-i\omega t}$, does not preclude the existence of *some* complete set of modes. Denote this set by φ_i. We may still formally define particle states by the requirement that the field decomposes as

$$\varphi = \Sigma_i(a_i\varphi_i + a_i^\dagger\varphi_i^*) \tag{2.18}$$

with $a_i|0>=0\ \forall i$.

Even then, however, problems persist as to how particles should be defined. These arise from the existence of other mode solutions, $\bar{\varphi}_j$ say, whereby we may alternatively decompose the field as

$$\varphi = \Sigma_i(\bar{a}_j\bar{\varphi}_j + \bar{a}_j^\dagger\bar{\varphi}_j^*) \tag{2.19}$$

with $\bar{a}_j|\bar{0}>=0\ \ \forall j$. (This situation exists also in the flat space case.) Then since both sets of mode solutions are complete, we can expand one set in terms of the other so that

$$\bar{\varphi}_j = \Sigma_i(\alpha_{ji}\varphi_i + \beta_{ji}\varphi_i^*) \tag{2.20}$$

and similarly for the φ_j modes. The relations between the two sets are known as a Bogolubov transformation. The (Bogolubov) coefficients, α_{ji} and β_{ji} are

found to be

$$\alpha_{ji} = (\bar{\varphi}_i, \varphi_j) \quad \beta_{ij} = -(\bar{\varphi}_i, \varphi_j^*) \tag{2.21}$$

where $(,)$ indicates the Klein-Gordon inner product[8]. It follows that

$$a_i = \Sigma_j(\alpha_{ji}\bar{a}_j + \beta_{ji}^*\bar{a}_j^\dagger \tag{2.22}$$

and similarly for $\bar{a}_j$. It follows that the Fock space constructions based on the two sets of modes are different (the exact meaning of this will be discussed shortly when we reformulate the theory, see section 2.2) so long as $\beta_{ji} \neq 0$. For example,

$$a_i|\bar{0}>= \Sigma_j\beta_{ji}^*|\bar{1}_j > \neq 0 \tag{2.23}$$

In fact, evaluation of the number operator for the unbarred modes, $N_i = a_i^\dagger a_i$, reveals that the vacuum for the barred modes, $|\bar{0}>$, contains unbarred particles! Explicitly, we have

$$< \bar{0}|N_i|\bar{0} >= \Sigma_j|\beta_{ji}|^2 \tag{2.24}$$

thus showing that the vacuum, $|\bar{0}\rangle$ contains $\Sigma_j|\beta_{ji}|^2$ particles of the φ type.

Thus we are led to the conclusion that even in flat spacetime, there exist a plethora of vacua. Grammatically, we are no longer bound to speak of *the* vacuum, but of (the choice of) *a* vacuum. The notion of a 'natural' vacuum in flat spacetime comes from the fact that such a vacuum is associated with inertial observers [12]. When we come to consider curved spacetimes, it will turn out that when a spacetime admits some measure of flatness, then we may speak of a natural vacuum; in that case, we will usually be considering spacetimes which are asymptotically static, for which a natural notion of particles then exists.

2.1.2 Fock spaces

Just as the structure defined in section 1.1 is a suitable state space for single particle quantum *mechanics*, when it comes to quantum *field* theory, we rely on the notion of Fock space to act as the state space for the quantum

field. When we consider a finite collection of (decoupled) osicllators with frequencies $\omega_1, \omega_2, \ldots, \omega_n$, we take the new Hilbert space to be the tensor product of the Hilbert spaces, $\mathcal{F}_1, \ldots, \mathcal{F}_n$ of the individual oscillator systems:

$$\mathcal{F} = \mathcal{F}_1 \otimes, \cdots, \mathcal{F}_n \cong L^2(Q) \tag{2.25}$$

where Q is the classical configuration space of the individual oscillator systems. The tensor product is defined as follows: for two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, let V denote the set of linear maps $A : \bar{\mathcal{H}}_1 \to \mathcal{H}_2$, such that the range of A is a finite dimensional subspace of $\mathcal{H}_2$. V has a natural vector space structure. We may define an inner product on V by

$$< A, B >_V = \mathrm{tr}(A^\dagger B). \tag{2.26}$$

The tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ is then the Hilbert space completion of V. Seen another way, $\mathcal{H}_1 \otimes \mathcal{H}_2$ consists of all bilinear maps $\alpha : \bar{\mathcal{H}}_1 \times \bar{\mathcal{H}}_2 \to \mathcal{C}$ such that

$$\Sigma_{i,j} \mid \alpha(\bar{e}_{1i}, \bar{e}_{2j}) \mid^2 < \infty. \tag{2.27}$$

By induction, this construction may be generalised to the case of finitely many Hilbert spaces, as required in 2.25 above. Thus we may see $\mathcal{H}_1 \otimes \cdots \mathcal{H}_n$ as consisting of all multilinear maps $\alpha : \bar{\mathcal{H}}_1 \times \cdots \times \bar{\mathcal{H}}_n \to \mathcal{C}$ satisfying

$$\Sigma \mid \alpha(\bar{e}_{1i_1}, \ldots, , \bar{e}_{ni_n}) \mid^2 < \infty \tag{2.28}$$

If $\mathcal{H}_1 = \cdots = \mathcal{H}_n (= \mathcal{H})$, then the symmetrised tensor product, $\otimes_s^n \mathcal{H}$, is defined to be the subspace of the n-fold tensor product space, $\otimes^n \mathcal{H}$, consisting of maps, α, which are totally symmetric in the n-variables.

Finally, we arrive at the definition of the Fock space: let $\mathcal{H}$ be a Hilbert space. The Fock space associated with $\mathcal{H}$ is defined to be

$$\mathcal{F}(\mathcal{H}) = \oplus_{n=0}^{\infty} (\otimes^n \mathcal{H}). \tag{2.29}$$

[Here $\oplus$ indicates the direct sum defined as follows: if $\{\mathcal{H}_\alpha\}$ is a set of Hilbert spaces, then their Cartesian product $\times_\alpha \mathcal{H}$ (note:not the tensor product) produces elements which consist of collections of vectors $\{\phi_\alpha\}$ with $\phi_\alpha \in \mathcal{H}$.

The subset $V \subset \times_\alpha \mathcal{H}_\alpha$ composed of elements for which only some of the $\phi_\alpha \neq 0$ is an inner product space. Then the Hilbert space completion of V, denoted $\oplus_\alpha \mathcal{H}_\alpha$, is the direct sum Hilbert space]. We take $\otimes^0 \mathcal{H} = \mathbb{C}$; then explictly, we have

$$\mathcal{F}(\mathcal{H}) = (\mathbb{C}) \oplus (\otimes^1 \mathcal{H}) \oplus (\mathcal{H} \otimes \mathcal{H}) \oplus \cdots . \tag{2.30}$$

We can then define the symmetric Fock space $\mathcal{F}_s(\mathcal{H})$ by

$$\mathcal{F}_s(\mathcal{H}) = \oplus_s^\infty(\mathcal{H}) \quad = (\mathbb{C}) \oplus (\mathcal{H}) \oplus (\mathcal{H} \otimes_s \mathcal{H}) \oplus \cdots \tag{2.31}$$

Now, every state in the Fock space $\mathcal{F}(\mathcal{H})$ has a direct physical interpretation in terms of the probabilities for finding various numbers of particles in the various particle states. (However, according to Haag's theorem [43], the Fock representation is not admissible for interacting or self-interacting quantum field theories. This indicates further difficulties with the particle interpretation. See also [39].) When use is made of a (symmetric) Fock sub-space, this represents the indistinguishability of elementary particles: an interchange of particles produces the same physical state. The choice of the symmetric tensor product is related to the fact that we consider bosonic fields, which have integer spin and thus obey symmetric statistics, via the spin- statistics theorem.

2.2 Reformulation

'Tis better to be that which we destroy,
than by destruction dwell in doubtful joy. Macbeth.

Having lambasted the standard approach to quantum field theory, it is now incumbent upon us to provide an alternative. This is achieved by adopting a more rigorous standpoint regarding the nature of the field operator. First, however, to attempt a reformulation of fields such that the sum

$$\varphi = \Sigma_i [\psi_i a_i + \psi_i^* a_i^\dagger]$$

makes sense, we have to return once more unto the Hilbert space concept.

2.2.1 Single particle case

Recall that in section 1.3 we examined how the symplectic structure may be useful in examining quantum theory from a new standpoint. We should like to use it here in a similar capacity. We identified the symplectic vector space $(\mathcal{M}, \sigma)$ with the symplectic (vector) space of solutions $(\mathcal{S}, \sigma)$ to the equation of motion. This space is very useful for our work here. We begin by complexifying $\mathcal{S}$, thereby obtaining a $2n-$complex dimensional vector space, $\mathcal{S}^{\mathbb{C}}$. $\sigma(y, \cdot)$ is extendible to this new space. The map $(,) : \mathcal{S} \times \mathcal{S} \to \mathbb{C}$ defined by

$$(y_1, y_2) = -i\sigma(\bar{y}_1, y_2). \tag{2.32}$$

would satisfy the properties of an inner product were it not for the fact that it is not positive-definite (which causes difficulties in obtaining a suitable norm). We overcome this problem by focussing attention on solutions of the form

$$q_i = \alpha_i e^{-i\omega t}, \tag{2.33}$$

i.e. solutions which oscillate with positive frequency. Then the map $(,)$ is positive definite on this space of solutions, $\mathcal{H}$, thereby making $\mathcal{H}$ into an $n-$dimensional complex Hilbert space.

We may associate with this construction a finite-dimensional Fock space, $\mathcal{F}_s(\mathcal{H})$ (as described earlier) upon which we may define Heisenberg operators $q'_{iH}(t) = \xi_i(t)a_i + \bar{\xi}_i(t)a_i^\dagger$ and $p'_{iH}(t) = dq_{iH}/dt$. Thus we have constructed an alternative set $(\mathcal{F}'; q'_i; p'_i; H')$ to the usual one $(\mathcal{F}; q_i; p_i; H)$ constructed in standard approaches (see section 2.1.2). But since we are dealing with finitely many oscillators, we may think of the two sets as equivalent in the sense defined by the following theorem:

Theorem 2.1 *(Stone-von Neumann): Let $(\mathcal{M}, \sigma)$ be a finite dimensional symplectic vector space. Let $(\mathcal{F}, \hat{W}(y))$ and $(\mathcal{F}', \hat{W}'(y))$ be strongly continuous, irreducible, unitary representations of the Weyl relations*

$$\hat{W}(y_1)\hat{W}(y_2) = e^{i\sigma(y_1, y_2)}\hat{W}(y_1 + y_2) \tag{2.34}$$

and

$$\hat{W}^\dagger(y) = \hat{W}(-y) \tag{2.35}$$

Then $(\mathcal{F}, \hat{W}(y))$ *and* $(\mathcal{F}', \hat{W}'(y))$ *are unitarily equivalent.*

[By unitary equivalence, we mean the following: a space $\mathcal{F}$ together with any collection of operators $V_\alpha : \mathcal{F} \to \mathcal{F}$ is unitarily equivalent to $(\mathcal{F}, \mathcal{V}'_\alpha)$ if there exists a unitary map $U : \mathcal{F} \to \mathcal{F}$ satisfying $U^{-1}V'_\alpha U = V_\alpha$, for all α. The Weyl relations above involve the observables $W(y) = e^{i\sigma}$, obtained by 'exponentiating' the fundamental observables σ.] Now the construcion of the Hilbert space above utilised the time-independence of the Hamiltonian, H'. If it is not time-independent, then solutions which oscillate with purely positive frequency will not exist. Then it is difficult, if not impossible, to construct a suitable subspace $\mathcal{H} \subset \mathcal{S}^{\mathbb{C}}$ which will play the role of the Hilbert space in the manner indicated above. Instead, we proceed as follows. Let $(\mathcal{S}, \sigma)$ be the symplectic vector space of solutions to the equations of motion with a time-dependent Hamiltonian. Complexify $\mathcal{S}$ as usual, and define a (non-positive definite) inner product on $\mathcal{S}^{\mathbb{C}}$ by

$$(y_1, y_2) = -i\sigma(\bar{y}_1, y_2) \tag{2.36}$$

again. However, there is now no natural way to decompose $\mathcal{S}^{\mathbb{C}}$ into a subspace of positive frequency solutions. This problem is circumvented by choosing any subspace $\mathcal{H} \subset \mathcal{S}^{\mathbb{C}}$ which satisfies the following properties:

1. The inner product is positive definite on $\mathcal{H}$, thus making it into a Hilbert space over $\mathbb{C}$.
2. $\mathcal{S}^{\mathbb{C}} = \mathcal{H} + \bar{\mathcal{H}}$
3. for all $z^+ \in \mathcal{H}, z^- \in \bar{\mathcal{H}}$, we have $(z^+, z^-) = 0$

These properties allow the decomposition of any $z \in \mathcal{S}^{\mathbb{C}}$ as $z = z^+ + z^-$. We obtain a real-linear, one-to-one, onto map $K : \mathcal{S} \to \mathcal{H}$, defined for all $y \in \mathcal{S}$ by $Ky = y^+$. We have

$$\begin{aligned} \mathrm{Im}(Ky_1, Ky_2)_{\mathcal{H}} &= -\mathrm{Re}(\sigma(\bar{K}y_1, Ky_2) \\ &= -\frac{1}{2}\sigma(\bar{K}y_1, Ky_2) - \frac{1}{2}(Ky_1, \bar{K}y_2) \\ &= -\frac{1}{2}\sigma(y_1, y_2) \end{aligned} \tag{2.37}$$

This last line follows from (iii) above and the result that for all $y \in \mathcal{S}^{\mathbb{C}}$, we have $y = Ky + \bar{K}y$.

Before leaving this section, we make a few remarks concerning this construction:

Remarks

- The relationship between the space of solutions and the Hilbert space is a manifestation of the so-called wave-particle duality: elements of $\mathcal{S}$ are wave solutions of the classical equations of motion, while elements of $\mathcal{H}$ are referred to as single-particle wavefunctions in the context of quantum mechanics.

- It should be evident that, in a sense, the inner product is a purely 'classical' matter, i.e reliant only on the underlying symplectic structures.

- The choice of subspace is equivalent to the specification of a real inner product $\mu : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ on the original space of solutions: for $y_1, y_2 \in \mathcal{S}$, define

$$\begin{aligned} \mu(y_1, y_2) &= Re(Ky_1, Ky_2)_{\mathcal{H}} \\ &= \mathrm{Im}\sigma(\bar{K}y_1, Ky_2) \end{aligned} \tag{2.38}$$

so that

$$(Ky_1, Ky_2)_{\mathcal{H}} = \mu(y_1, y_2) - \frac{1}{2}\sigma(y_1, y_2) \tag{2.39}$$

Now for all $z_1, z_2 \in \mathcal{H}$, we have $\| z_1 \|^2 \| z_2 \|^2 \geq |(z_1, z_2|^2 \geq |\mathrm{Im}(z_1, z_2|^2$ Taking $z_1 = Ky_1$, $z_2 = Ky_2$, it follows that for all $y_1, y_2 \in \mathcal{S}$

$$\mu(y_1, y_1)\mu(y_2, y_2) \geq \frac{1}{4}[\sigma(y_1 y_2]^2 \tag{2.40}$$

This result may be strengthened (since the Schwarz inequality may always be 'saturated', i.e. there exists some z_1, z_2 such that $\| z_1 \|^2 \| z_2 \|^2 = |(z_1, z_2|^2)$ to give

$$\mu(y_1, y_1) = \frac{1}{4} max_{y_2 \neq 0} \frac{[\sigma(y_1 y_2]^2}{\mu(y_2, y_2)}$$

This completes our reformulation of the single particle case.

2.2.2 Fields

We now exhibit a reformulation of the field theory case analogous to the work we have just done for the single particle case. Some of the mathematical difficulties associated with quantum field theory which we alluded to earlier may be cured by realising that the field, φ, is not an (operator-valued) *function*, but a distribution. At this point, the uninitiated may well sympathise with a certain famous mathematician who expressed his views thus on the topic[1]:

> *I turn away with fear and horror from this lamentable plague of functions which do not have derivatives.*
> Hermite, in a letter to Stieltjes.

Nowadays, however, the theory of distributions is well established. For us, the use of distributions means that the field, φ, is no longer defined on $\mathbb{R}^4$, but on $C_0^\infty(\mathbb{R}^4)$ (the space of smooth functions with compact support) [or some other suitable space of test functions, such as Schwartz space, $\mathcal{S}(\mathbb{R}^4)$]. A (real) distribution, D, on spacetime is simply a continuous, linear, real-valued map, $D : \mathcal{C}_o^\infty(\mathbb{R}^4) \rightarrow \mathbb{R}$. We thus replace the idea of $\varphi(x)$ as "the value of the field at x", and adopt instead the heuristic view of $\varphi(f)$ as the spacetime average of the (now ill-defined) $\varphi(x)$ with some averaging function f. $\varphi(f)$ is thus a smeared field, symbolically written as

$$\varphi(f) = \int_{\mathbb{R}^4} \varphi(x) f(x) d^4x. \tag{2.41}$$

(Not all distributions arise in this manner; one familiar example is the Dirac "δ - function".) From a quantum mechanical point of view, this smearing is rather appropriate, being in the spirit of the Uncertainty Principle. In particular Bohr and Rosenfeld (in [55]) used quantum laws to show that it is impossible to measure the electric field strength at a point.

Thus the idea of using distributions is well motivated from both a mathematical and physical standpoint. To see how they arise, we first examine the symplectic structure underlying field theory. We take our phase space, Γ,

[1] quoted in [35]

to consist of initial data which are smooth and of compact support on some three-dimensional hypersurface, Σ_0, i.e.

$$\Gamma \stackrel{def}{=} \{[\varphi, \pi] | \varphi, \pi : \Sigma_0 \rightarrow I\!R; \varphi, \pi \in \mathcal{C}_o^\infty(\Sigma_0)\}. \tag{2.42}$$

Every point $[\varphi, \pi] \in \Gamma$ uniquely determines a solution (which is an element of the space of solutions, $\mathcal{S}$) of the Klein-Gordon equation. On Γ we have the symplectic structure $\sigma :\rightarrow \Gamma \times \Gamma \rightarrow I\!R$ given by

$$\sigma\{[\varphi_1, \pi_1], [\varphi_2, \pi_2]\} = \int_{\Sigma_0} d^3x(\pi_1\varphi_2 - \pi_2\varphi_1). \tag{2.43}$$

The fundamental Poisson bracket relations on Γ are:

$$\{\sigma([\varphi_1, \pi_1], \cdot), \sigma([\varphi_2, \pi_2], \cdot)\} = -\sigma([\varphi_1, \pi_1], [\varphi_2, \pi_2]). \tag{2.44}$$

If we take $[\varphi_1, \pi_1] = [0, f_1]$ and $[\varphi_2, \pi_2] = [f_2, 0]$, then we obtain

$$\{\int f_1\varphi, \int f_2\pi\} = \int f_1 f_2 \tag{2.45}$$

using 2.44 above. The above relation is more loosely written as

$$\{\varphi(x_1), \pi(x_2)\} = \delta(x_1 - x_2). \tag{2.46}$$

To make the transition to quantum theory, we represent the functions $\sigma([\varphi_1, \pi_1], \cdot)$ on Γ, by operators $\hat{\sigma}([\varphi_1, \pi_1], \cdot)$ (acting on some suitable space of states), satisfying the commutation relations

$$[\hat{\sigma}([\varphi_1, \pi_1], \cdot), \hat{\sigma}([\varphi_2, \pi_2], \cdot)] = -i\sigma([\varphi_1, \pi_1], [\varphi_2, \pi_2])I \tag{2.47}$$

Note that the left hand side involves quantum objects, while the right hand side is just the classical function, σ. Because of the correspondence between Γ and $\mathcal{S}$, the space of solutions, we can define operators $\hat{\sigma}(\phi, \cdot)$ satisfying analogues of 2.47 above $\forall \phi \in \mathcal{S}$. Then given any $\phi \in \mathcal{S}$, we decompose it into its positive and negative frequency parts: $\phi = \phi^+ + \phi^-$. Let $\mathcal{S}^{\mathcal{C}+}$ denote the subspace spanned by the positive frequency parts of solutions in $\mathcal{S}$. Then define on $\mathcal{S}^{\mathcal{C}+}$ the Klein-Gordon inner product by

$$(\phi^+, \xi^+) = -i\sigma(\bar{\phi}^+, \xi^+) \tag{2.48}$$

Then Cauchy complete $\mathcal{S}^{\mathcal{C}}{}_{+}$ in the norm defined by this inner product to obtain a complex Hilbert space, $\mathcal{H}$. This $\mathcal{H}$ has been obtained without putting the system in a box (which generates the plane wave basis). The association $\phi \rightarrow \phi^{+}$ is equivalent to defining a real-linear one-to-one map $K : \mathcal{S} \rightarrow \mathcal{H}$ which takes $\mathcal{S}$ into a dense subspace of $\mathcal{H}$.

For the quantum theory (i.e the transition from classical to quantum), we take the Hilbert space to be $\mathcal{F}_s$. For each $\phi \in \mathcal{S}$, we define the operator $\hat{\sigma}(\phi, \cdot)$ on $\mathcal{F}_s$ by

$$\hat{\sigma}(\phi, \cdot) = ia(\bar{K}\phi) - i\alpha^{\dagger}(K\phi). \tag{2.49}$$

Then the Heisenberg representative of this operator is

$$\hat{\sigma}_H(\phi, \cdot) = ia(\bar{K}\phi_t) - ia^{\dagger}(K\phi_t), \tag{2.50}$$

where ϕ_t is the solution whose initial data at time t equals the initial data for ϕ at $t = 0$.

Using this, we may recover the 'field in a box' formalism from a more rigorous standpoint. This is done at the cost of losing the 'natural' choice of Hilbert space as determined by the plane wave basis of the former. We still have the apparently arbitrary choice of a suitable subspace of $\mathcal{S}$ to deal with. This choice is, however, not irrevelant in field theory. Earlier we saw that in the case of finitely many oscillators that via the Stone-von Neumann theorem, the choice of $\mathcal{H}$ did not affect measurements as such, since the different choices led to unitarily equivalent constructions. In field theory, however, this is not necessarily the case. Our choice of 'natural' in the specification of $\mathcal{H}$ followed from the time translation invariance of the classical theory (which is an aspect of the Poincaré invariance). When we attempt to do quantum field theory in curved spacetimes, we will not have such luxury. Thus we will not have a case of 'natural selection' of an $\mathcal{H}$, and so have to deal with the large class of unitarily inequivalent constructions arising from the freedom in choosing $\mathcal{H}$. The question of how to appreciate and use unitarily equivalent constructions is dealt with in subsequent sections. This is important in various cases, even in flat spacetime (where the 'natural' choice is not the only one available), and stationary spacetimes.

Before doing so, we first deal with the aspects of the choice of $\mathcal{H}$ which are intrinsic to the construction made above (and thus not related to specific aspects such as time translation invariance). This will allow us to deal more effectively with curved spacetime scenarios. In field theory, the choice of $\mathcal{H}$ cannot be simplified to the three steps shown earlier for the case of finitely many oscillators (see 2.2.1).This is due to the fact that $\mathcal{S}$ is now infinite dimensional: we first need to Cauchy-complete it before $\mathcal{H}$ can be viewed as a subspace of $\mathcal{S}^{\mathbb{C}}$. In fact it is a rather disturbing result that no non-zero element of $\mathcal{H}$ actually lies in $\mathcal{S}^{\mathbb{C}}$ before this completion. But, this enlargement is really part of the specification of $\mathcal{H}$ itself! So we seek a construction which both completes/enlarges $\mathcal{S}^{\mathbb{C}}$ and specifies $\mathcal{H}$. This may be accomplished with the introduction of a *complex structure*, $J : \mathcal{S}_\mu \to \mathcal{S}_\mu$ with $J^2 = -1$ where $\mathcal{S}_\mu$ is the space obtained by enlarging $\mathcal{S}$ using a real inner product, μ. We refer the reader to the literature for the details [6],[54].

This construction results in the complexification of $\mathcal{S}$ to $\mathcal{S}_\mu^{\mathbb{C}}$. We may then define $K : \mathcal{S}_\mu^{\mathbb{C}} \to \mathcal{H}$ to be the orthogonal projection map onto the subspace $\mathcal{H}$ of $\mathcal{S}_\mu^{\mathbb{C}}$. The restriction of K to $\mathcal{S}$ then defines a real-linear map $K : \mathcal{S} \to \mathcal{H}$ such that for all $\phi_1, \phi_2 \in \mathcal{S}$, we have

$$\begin{aligned}(K\phi_1, K\phi_2)_{\mathcal{H}} &= -i\sigma(\bar{K\phi_1}, K\phi_2) \\ &= \mu(\phi_1, \phi_2) - \frac{i}{2}\sigma(\phi_1, \phi_2) \end{aligned} \tag{2.51}$$

Thus

$$Im(K\phi_1, K\phi_2)_{\mathcal{H}} = -\frac{1}{2}\sigma(\phi_1, \phi_2). \tag{2.52}$$

We are now in a position to demonstrate the manner in which the field, φ is interpreted as an "operator-valued distribution". This arises from the relationship between $\mathcal{S}$ and the vector space $\mathcal{D}(R^4)$ of test functions on Minkowski spacetime which we now exhibit. Now $(\mathcal{S}, \sigma)$ is the symplectic vector space of solutions to the Klein- Gordon equation with initial data $[\varphi, \pi]$ lying in $C_0^\infty(\Sigma_0)$. Let $f \in \mathcal{D}(R^4)$ and let Af and Rf denote the advanced and retarded solutions to the Klein-Gordon equation with source f, respectively (thus they are not elements of $(\mathcal{S}, \sigma)$). So:

$$(\partial^a \partial_a - m^2)(Af) = f \tag{2.53}$$

$$(\partial^a \partial_a - m^2)(Rf) = f \tag{2.54}$$

with $Af = 0$ outside the causal past of the support of f and $Rf = 0$ outside the causal future of the support of f. Then

$$Ef = Af - Rf \tag{2.55}$$

satisfies the (homogeneous) Klein-Gordon equation

$$(\partial^a \partial_a - m^2)Ef = 0. \tag{2.56}$$

Thus Ef is an element of $\mathcal{S}$, and in fact, E may be viewed as a linear map $E : \mathcal{D}(\mathbb{R}^4) \to \mathcal{S}$ (since the initial data for the solution lie in $C_0^\infty(\Sigma_0)$, and $f \in \mathcal{D}(\mathbb{R}^4)$). (We will examine in some detail the existence of the solutions $Af, Rf \text{and} Ef$ in section 3.4.1). Some important properties of this map are contained in the following result:

Lemma 2.2 *(1) E is onto, i.e. every $\phi \in \mathcal{S}$ can be expressed as $\phi = Ef$ for some $f \in \mathcal{T}$.*
(2) $Ef = 0 \Leftrightarrow f = (\partial^a \partial_a - m^2)g$ for some $g \in \mathcal{D}(\mathbb{R}^4)$.
(3) $\forall \phi \in \mathcal{S}$ and $\forall f \in \mathcal{D}(\mathbb{R}^4)$, we have

$$\int \sigma f d^4 = \sigma(Ef, \phi) \tag{2.57}$$

This last property shows that for each $f \in \mathcal{D}(\mathbb{R}^4)$, the function $\sigma(Ef, \cdot)$ on $\mathcal{S}$ is equal to what is obtained by averaging the solution over spacetime, weighted by f. Thus we identify the (Schrödinger) operator $\sigma(Ef, \cdot)$ defined by eqn.(2.57) above with the spacetime average of the (Heisenberg) quantum field operator, weighted by f. That is to say, we <u>define</u> the operator $\hat{\varphi}(f)$ by

$$\begin{aligned} \hat{\varphi}(f) &= \hat{\sigma}(Ef, \cdot) \\ &= ia(K(\bar{E}f)) - ia^\dagger(K(Ef)). \end{aligned} \tag{2.58}$$

This provides us with a mathematically rigorous definition of the Heisenberg field operator $\hat{\varphi}$ as an operator-valued distribution on spacetime, i.e. as a map from test functions into operators on $\mathcal{F}_s$. The collection of all "smeared field

operators" of the form above comprise the fundamental observables of the theory (we emphasise again that they are fundamental because, like the q_μ and p_μ of the classical case, all other observables may be derived from them). Now $\hat{\varphi}$ satisfies the distributional version of the Klein-Gordon equation:

$$\hat{\varphi}([\partial^a\partial_a - m^2]g) = \sigma(E[\partial^a\partial_a - m^2]g, \cdot) = 0. \tag{2.59}$$

The fundamental commutation relations are

$$\begin{aligned} [\hat{\varphi}(f), \hat{\varphi}(g)] &= [\sigma(Ef, \cdot), \sigma(Eg, \cdot)] \\ &= -i\sigma(Ef, Eg) \\ &= -iE(f, g) \end{aligned} \tag{2.60}$$

with $E(f,g) = \int fEgd^4x$. This relation is more loosely expressed as

$$[\hat{\varphi}(x), \hat{\varphi}(x')] = -iE(x, x'). \tag{2.61}$$

With the use of eqn.(2.58), we can write

$$\begin{aligned} < 0|\hat{\varphi}(f)\hat{\varphi}(g)|0 > &= (KEf, KEg)_{\mathcal{H}} \\ &= \mu(Ef, Eg) - \frac{i}{2}E(f, g) \end{aligned} \tag{2.62}$$

which shows that the bilinear map $\mu : \mathcal{S} \times \mathcal{S} \to I\!R$ is just the real part of the two-point function $< 0|\hat{\varphi}(x)\hat{\varphi}(x')|0 >$ of the quantum field (in the vacuum state defined exactly by the choice of μ).

2.3 Unitary Equivalence, a.k.a. the S-matrix.

Recall that for systems with finitely many degrees of freedom, the different constructions will always be unitarily equivalent, by the Stone- von Neumann theorem. The case of fields, on the other hand, will generally bring into consideration unitarily inequivalent constructions. But it is still possible to analyse cases where different constructions result in equivalent theories. In such scenarios, we may construct the familiar S- matrix. The importance of unitarily equivalent constructions stems from the following consideration:

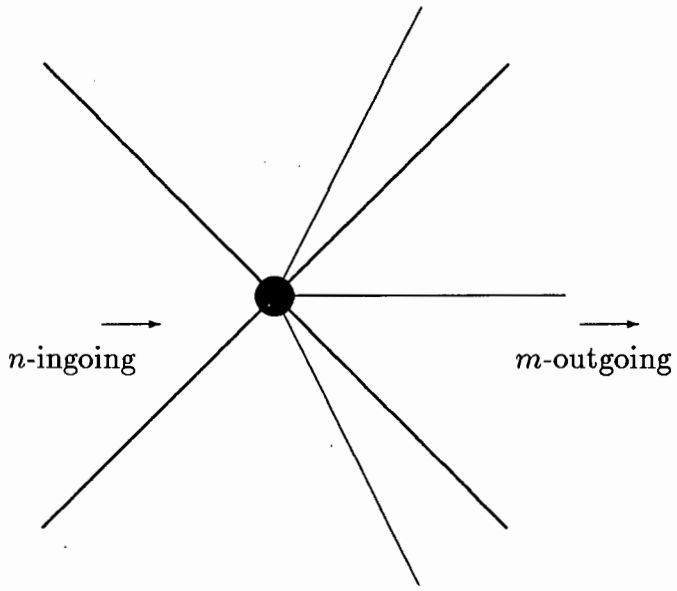

Figure 2.1: The usual view of the S-matrix: n particles enter, and m particles leave.

we saw in section 2.1.1 that a useful concept of particle persists in curved spacetimes which are asymptotically static. Now if a spacetime possesses two such regions (say an 'in' and an 'out' region), then we may construct accompanying Fock spaces, $\mathcal{F}_s(\mathcal{H}_{in})$ and $\mathcal{F}_s(\mathcal{H}_{out})$, respectively. If we are able to exhibit unitary equivalence between these two constructions, then we may 'compare' them; specifically, we may examine what the vacuum vector $|0_{in}\rangle \in \mathcal{F}_s(\mathcal{H}_{in})$ 'looks like' in $\mathcal{F}_s(\mathcal{H}_{out})$, particularly in terms of particle number in the latter. It would help to remember here that the vector denoted $|0_{in}\rangle$ is just an element of a Hilbert space, picked out to be a 'vacuum' by the choice of basis, so that a change of basis may bring about a re-interpretation. Our derivation of the S-matrix is non-standard, and is here viewed as a unitary map between Hilbert spaces. It would help to keep the figure below in mind.

The results we shall outline hold for any symplectic vector space, $(\mathcal{S}, \sigma)$. We shall (later) apply it in particular to the case where $\mathcal{S}$ is the space of

solutions to the Klein-Gordon equation in a globally hyperbolic spacetime. This work follows closely that of [51].

Let $\mu_1 : \mathcal{S} \times \mathcal{S} \to I\!R$ and $\mu_2 : \mathcal{S} \times \mathcal{S} \to I\!R$ be two bilinear maps satisfying 2.40 above. We can thus construct two Fock spaces, $\mathcal{F}_1$ and $\mathcal{F}_2$ with operators $\hat{\sigma}_1(\phi, \cdot) : \mathcal{F}_1 \to \mathcal{F}_2$ and $\hat{\sigma}_2(\phi, \cdot) : \mathcal{F}_2 \to \mathcal{F}_2$. Our interest lies in analysing the conditions under which there exists a unitary map $U : \mathcal{F}_1 \to \mathcal{F}_2$ such that for all $\phi \in \mathcal{S}$, we have

$$U\hat{\sigma}_1(\phi, \cdot)U^{-1} = \hat{\sigma}_2(\phi, \cdot) \tag{2.63}$$

The analysis of unitary equivalence divides into two cases:

Case I: There exist $C, C' > 0$ such that

$$C'\mu_1(\phi, \phi) \geq \mu_2(\phi, \phi) \geq C\mu_1(\phi, \phi) \tag{2.64}$$

$\forall \phi \in \mathcal{S}$.

Case II: No such $C, C' > 0$ exist.

It can be proven that unitary equivalence cannot hold in Case II [54]. Thus we focus attention on Case I.

Eqn. 2.64 shows that μ_1 and μ_2 define 'equivalent norms' on $\mathcal{S}$. They define identical Cauchy completions of $\mathcal{S}$, denoted by $\mathcal{S}_\mu$. The two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, associated with the maps μ_1 and μ_2 respectively, are thus different subspaces of the same space, viz. $\mathcal{S}_\mu^{\mathbb{C}}$. As usual, we can define an inner product, $2\mu_1(\bar{\phi}, \xi)$ on $\mathcal{S}_\mu^{\mathbb{C}}$, as well as the orthogonal projection map $K_1 : \mathcal{S}_\mu^{\mathbb{C}} \to \mathcal{H}_1$. Similarly, let $\bar{K}_1 : \mathcal{S}_\mu^{\mathbb{C}} \to \bar{\mathcal{H}}_1$ denote the orthogonal projection map onto $\bar{\mathcal{H}}_1$. Since $\mathcal{H}_1$ and $\bar{\mathcal{H}}_1$ are orthogonal subspaces which span $\mathcal{S}_\mu^{\mathbb{C}}$, we have $K_1 + \bar{K}_1 = I$ on $\mathcal{S}_\mu^{\mathbb{C}}$. Let $K_2 : \mathcal{S}_\mu^{\mathbb{C}} \to \mathcal{H}_2$ and $\bar{K}_2 : \mathcal{S}_\mu^{\mathbb{C}} \to \bar{\mathcal{H}}_2$ denote the projection maps for the inner product $2\mu_2(\bar{\phi}, \xi)$ on $\mathcal{S}_\mu^{\mathbb{C}}$

By using 2.63 above, we see that the two constructions are unitarily equivalent if and only if there exists a unitary map $U : \mathcal{F}_1 \to \mathcal{F}_2$ such that

$$U[ia_1(\bar{K}_1\phi) - ia^\dagger(K_1\phi)]U^{-1} = ia_2(\bar{K}_2\phi) - ia^\dagger(K_2\phi) \tag{2.65}$$

for all $\phi \in \mathcal{S}$. The equation above also holds for all $\phi \in \mathcal{S}_\mu^{\mathbb{C}}$.

Let $A : \mathcal{H}_2 \to \mathcal{H}_1$ and $B : \mathcal{H}_2 \to \bar{\mathcal{H}}_1$ denote the restrictions of K_1 and $\bar{K}_1$, respectively, to the subspace $\mathcal{H}_2 \subset \mathcal{S}_\mu^{\mathcal{T}}$. Similarly, let $C : \mathcal{H}_1 \to \mathcal{H}_2$ and $D : \mathcal{H}_1 \to \bar{\mathcal{H}}_2$ denote the restrictions of K_2 and $\bar{K}_2$, respectively to $\mathcal{H}_1$. Then choosing $\phi \in \bar{\mathcal{H}}_1$, and writing $\chi = \bar{\phi}$, we see that unitary equivalence implies that for all $\chi \in \mathcal{H}_1$, we have

$$U a_1(\bar{\chi}) U^{-1} = a_2(C\bar{\chi}) - a_2^\dagger(D\bar{\chi}). \tag{2.66}$$

These operators A, B, C, D are bounded and satisfy some important properties; they will carry us through the rest of our discussion on unitary equivalence.

Suppose we took $\chi, \phi \in \mathcal{H}_2$; then

$$\begin{aligned}(\phi, \chi) &= -i\sigma(\bar{\phi}, \chi) \\ &= -i\sigma(\overline{K_1\phi + \bar{K}_1\phi}, K_1\chi + \bar{K}_1\chi) \\ &= (A\phi, A\chi)_{\mathcal{H}_1} - (B\phi, B\chi)_{\bar{\mathcal{H}}_1}\end{aligned}$$

It follows that

$$A^\dagger A - B^\dagger B = I. \tag{2.67}$$

One can also show that

$$A^\dagger \bar{B} = B^\dagger \bar{A}, \tag{2.68}$$

where the bar denotes the corresponding map on the complex conjugate spaces. We also obtain similar formulae for the operators C and D:

$$\begin{aligned}C^\dagger C - D^\dagger D &= I \\ C^\dagger \bar{D} &= D^\dagger \bar{C}.\end{aligned} \tag{2.69}$$

In addition, for $\phi \in \mathcal{H}_1, \chi \in \mathcal{H}_2$, we get

$$\begin{aligned}(\phi, A\chi)_{\mathcal{H}1} &= -i\sigma(\bar{\phi}, K_1\chi) \\ &= -i\sigma(\bar{\phi}, K_1\chi + \bar{K}_1\chi) \\ &= -i\sigma(\bar{\phi}, \chi) \\ &= -i\sigma(\overline{K_2\phi + \bar{K}_2\phi}, \chi) \\ &= -i\sigma(\overline{K_2\phi}, \chi) \\ &= (C\phi, \xi)\end{aligned}$$

Thus we obtain

$$A^\dagger = C \tag{2.70}$$

Similar calculations reveal that

$$\bar{B}^\dagger = -D \tag{2.71}$$

These are all the conditions we need to satisfactorily define unitary equivalence. A unitary map U satisfying 2.66, with operators A, B, C, D satisfying 2.67 to 2.69, is known as a Bogolubov transformation. Assuming that such a map exists, we wish to calculate its effect on the vector $|0> \in \mathcal{F}_1$: Thus we let

$$\Psi = U|0> \tag{2.72}$$

and attempt to solve for the vector $\Psi \in \mathcal{F}_2$. We write Ψ as

$$\Psi = c(1, \phi^a, \phi^{ab}, \ldots) \tag{2.73}$$

(see section 2.1.2). We use the operators as follows: let $\xi \in \mathcal{H}_2$, and let $\chi = C^{-1}\xi$; then apply 2.66 to Ψ. We thus obtain:

$$Ua_1(\overline{C^{-1}\xi})U^{-1}\Psi = a_2(\overline{CC^{-1}\xi}) - a_2^\dagger(\overline{DC\xi}). \tag{2.74}$$

But using the definition 2.73, the left hand side vanishes; we then get

$$0 = [a_2(\bar{\xi}) - a_2^\dagger(\mathcal{E}\bar{\xi})]\Psi \tag{2.75}$$

where $\mathcal{E} : \bar{\mathcal{H}}_2 :\rightarrow \mathcal{H}_2$ is the symmetric operator (i.e. $\bar{\mathcal{E}}^\dagger = \mathcal{E}$) defined by

$$\mathcal{E} = \bar{D}\bar{C}^{-1} \tag{2.76}$$

Equation (2.75) above allows us to write for the 'components' of Ψ:

$$\bar{\xi}_a\phi^a = 0 \tag{2.77}$$

$$\sqrt{2}\bar{\xi}_a\phi^{ab} = 0 \tag{2.78}$$

$$\sqrt{3}\bar{\xi}_a\phi^{abc} = \sqrt{2}(\mathcal{E}\bar{\xi})^{(b}\phi^{c)} \tag{2.79}$$

$$\sqrt{4}\bar{\xi}_a\phi^{abcd} = \sqrt{3}(\mathcal{E}\bar{\xi})^{(b}\phi^{cd)} \tag{2.80}$$

The solution of eqn.(2.77) is $\phi^a = 0$. By induction, all the n-particle amplitudes vanish for odd n. For 2.78, we get that $\phi^{ab} = \mathcal{E}/\sqrt{2}$ (as maps). This implies that

$$\mathrm{tr}(\mathcal{E}^\dagger\mathcal{E}) < \infty. \tag{2.81}$$

This is equivalent to the conditions $\mathrm{tr}(D^\dagger D) < \infty$, and $\mathrm{tr}(B^\dagger B) < \infty$. In terms of the maps μ_1, μ_2, these conditions mean that the linear map Q : $\mathcal{S}_\mu \to \mathcal{S}_\mu$ defined by $\mu_1(\phi_1, Q\phi_2) = \mu_2(\phi_1, \phi_2) - \mu_1(\phi_1, \phi_2)$ be of trace class. Finally, the solution for Ψ may be written as

$$\Psi = c(1, 0, \sqrt{\frac{1}{2}}\varepsilon^{ab}, 0, \sqrt{\frac{3\cdot 1}{4\cdot 2}}\varepsilon^{(ab}\varepsilon^{cd)}, 0, \ldots) \tag{2.82}$$

where ε^{ab} denotes the "2-particle state" element of $\mathcal{H}_2 \otimes \mathcal{H}_2$ corresponding to the map $\mathcal{E}$. As for states other than the vacuum, such as a one-particle state, $a_1^\dagger(\chi)|0>$ for example, we can obtain this unitary action by applying the adjoint of 2.66 to $\Psi = U|0>$. The equation giving us the result of the unitary map on $|0>$ (viz. 2.82) tells us straightaway that particles are always created in pairs, and perhaps more importantly, that there will be a non-zero amplitude for particle creation $\Longleftrightarrow$ some initially purely positive frequency classical solution picks up a non-zero negative frequency part in the asymptotic future (since then $D \neq 0$ and hence $\mathcal{E} \neq 0$).

Chapter 3

Quantum fields in curved spacetimes

Out of the frying pan and into the fire... Anon.

Our aim in this chapter is an exposition of some fundamental results which follow from an analysis of quantum fields in a curved spacetime. A general justification for examining the effects of curvature on quantum fields may be found in considerations which follow from our remarks in section 2.1.1. We saw there that the flaws of the particle concept were exposed via the lack of a natural basis (of the appropriate Hilbert space) in non-Minkowski spacetimes. (Even in Minkowski spacetime, of course, the 'natural' nature of the basis was shown to be observer-related). We further interpret this to mean that a further investigation of the effects of curvature on quantum theory will bring to the fore other flaws in its foundations. In a related way, Penrose[33] has proposed that gravitational effects in quantum gravity might help to explain the mysterious 'collapse of the wavefunction' scenario. Now, besides the interest in the possible existence of flaws in the formulation of quantum fields in curved spacetimes, we will be interested in the propagation of such fields. More specifically, we will want to examine what happens under certain conditions when an initially 'vacuum' field interacts with a collapsing gravitational body. Hawking's celebrated result[25] showed that such a field will end up in a many-particle state. It turns out[54] that this mechanism

requires the existence of a bifurcate Killing horizons. This we now review. (Other geometric concepts will be found in appendix A.)

3.1 Geometrical Preliminaries

Let $(\mathcal{M}, g_{ab})$ be a spacetime possessing a one-parameter group of isometries generated by a Killing field, χ^a. A *Killing horizon* in $\mathcal{M}$ is a null surface, H, to which the Killing field χ^a is normal [53]. This implies that χ^a is tangent to the null geodesic generators of H. (We must keep in mind here that the Killing vector χ^a is null, so is perfectly capable of being both normal to H and tangent to the generators of H.) Since $\chi^a\chi_a = 0$ on H, we have that $\nabla^a(\chi^b\chi_b)$ also is normal to H, and thus parallel to χ^a. The *surface gravity*, κ, on H is defined by this relation, i.e.

$$\nabla^a(\chi^b\chi_b) = -2\kappa\chi^a \tag{3.1}$$

It turns out that κ is constant along each null geodesic generator of H. Killing's equation

$$\nabla_{(a}\chi_{b)} = 0 \tag{3.2}$$

then implies that

$$\chi^b\nabla_b\chi^a = \kappa\chi^a \tag{3.3}$$

This result shows that κ measures the failure of "Killing parameter time", v, to coincide with "affine parameter time", V, on H. That this is so may be seen by looking at the left hand side of the above equation as the gradient with respect to affine parameter of the Killing vector. This gradient, if zero, would demonstrate that the Killing vector is parallel transported with respect to affine parameter. But the fact is that the gradient turns out to be proportional to the vector, with "proportionality constant", κ.

Furthermore, the equation above can be solved to yield the relationship between the two parameters:

$$V = e^{\kappa v} \tag{3.4}$$

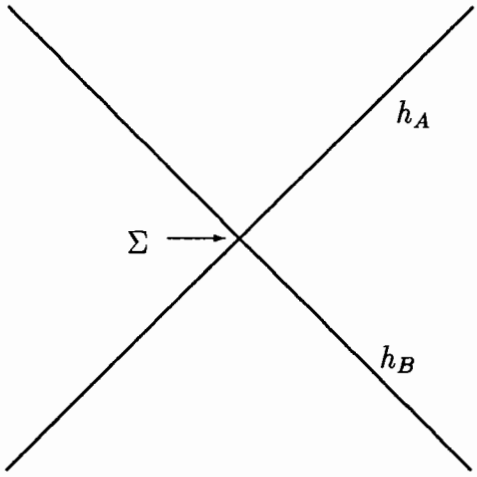

Figure 3.1: A bifurcate Killing horizon, consisting of the null surfaces h_A and h_B, which intersect at the two dimensional (spacelike) surface, Σ.

i.e. if v is a Killing parameter along the generators of H, then V is an affine parameter along the generators. (Heuristically, the simple exponential nature of the solution follows from the fact that the gradient of the vector is proportional to the vector itself!) In addition, we have that on H, the following relationship holds:

$$\chi^a = (\frac{\partial}{\partial v})^a = \kappa V(\frac{\partial}{\partial V})^a \tag{3.5}$$

Consequently, if $\kappa \neq 0$ and the null generators of H are geodesically complete, we must have $\chi^a = 0$ along each generator of H, i.e. χ^a vanishes on a spacelike cross-section, Σ, of H. This implies that if $\kappa \neq 0$, then H has the structure of a *bifurcate Killing horizon*, as shown in figure 3.1.

We are particularly interested in the relationship between the event horizons of black holes and Killing horizons. In this regard, we conclude this section with an important result:

Theorem 3.1 *Let $(\mathcal{M}, g_{ab})$ be a stationary, asymptotically flat spacetime*

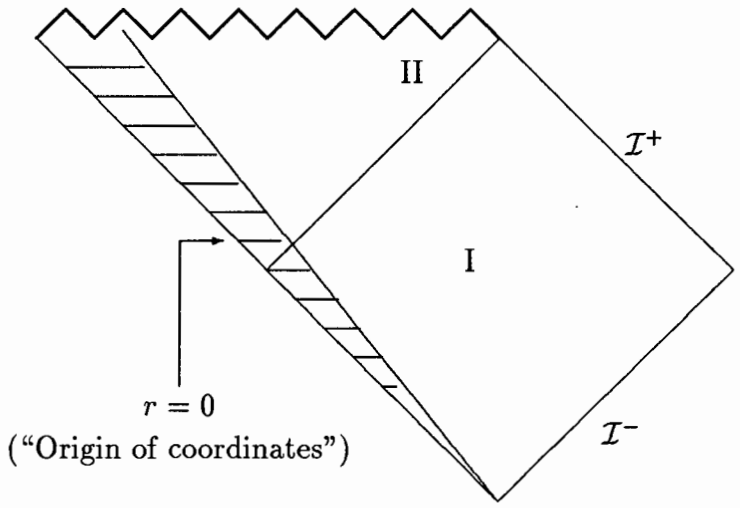

Figure 3.2: A conformal diagram of a spacetime in which a spherical body undergoes collapse to a Schwarzschild black hole.

containing a black hole, which is a solution of Einstein's equation with matter satisfying suitable hyperbolic equations. Then the event horizon, h^+, is a Killing horizon.

3.2 Particle Creation by Black Holes

The Hawking effect may be derived by a number of methods([25],[48],[17]); we shall follow that of [54]. Essentially, we are interested in the decomposition of some quantum field before and after its interaction with the collapsing gravitational body (see figure(3.2)) (In this way the particle creation effect by black hole is much like any other external field theory[51].) We assume that the body is asymptotically static in the past, i.e. before the collapse. We can construct in the familiar manner the quantum field theory and associated Fock space, $\mathcal{F}_s(\mathcal{H}_{in})$. Thus we have a natural particle interpretation of this Fock space in terms of particles in the asymptotic past.

As for the 'out' field, we have to be a bit more careful. This is due to the failure of region I (analogous to region I in extended Schwarzschild spacetime) to admit a Cauchy surface for the whole spacetime in the future.[1] (Despite this, the spacetime is static in region I in the asymptotic future). We then examine the symplectic space of solutions, $\mathcal{S}$; specifically,
I: we shall assume that each $\phi \in \mathcal{S}$ either reaches infinity, ($\mathcal{I}^+$) and/or the black hole horizon, h_A.
II:we seek a decomposition of $\mathcal{S}$ as $\mathcal{S} = \mathcal{S}_E \oplus \mathcal{S}_L$.

This decomposition will reflect whether the solution $\phi \in \mathcal{S}$ reached its destination (*i.e.* either $\mathcal{I}^+$ or h_A) at early or late times. We require that $\mathcal{S}_L$ be so defined that the spacetime is static when its elements reach $\mathcal{I}^+$ or h_A. This can be achieved by choosing a Cauchy surface, Σ, which intersects the horizon outside the collapsing matter. Why would this work? Well, we can then take $\underline{\mathcal{S}_E}$ to be those elements of $\mathcal{S}$ which have support *inside* the black hole at the 'time' defined by Σ (thus they have reached their destination early, in this case the destination being h_A); then $\underline{\mathcal{S}_L}$ consists of those solutions which lie *outside* the black hole.

Since the Hawking effect is independent of the details of the collapse, we are led to the notion that the use of the maximally extended Schwarzschild (vacuum) solution will be of greater value than the collapsing body spacetime. (This means that the effect is more a consequence of the causal and topological structure of spacetime than the specific geometry.)

We may thus naturally set up a correspondence between $\mathcal{S}_L$ and the solutions, $\mathcal{S}^S$ in extended Schwarzschild spacetime with initial data on a corresponding Cauchy surface $\bar{\Sigma}$ in the latter. Thus we may identify $\mathcal{S}_L \ni \phi_L$ with $\tilde{\phi} \in \mathcal{S}^S$. We use the latter solutions to define the out representation: $\forall \phi_L \in \mathcal{S}_L$, let $K\phi_L \in \mathcal{H}_{out}$ be the solution which corresponds to the positive frequency part of $\tilde{\phi}$ in extended Schwarzschild spacetime. (Of course, $K\mathcal{S}_L \neq \mathcal{H}_{out}$: we have not yet dealt with $\mathcal{S}_E$.)

Let $\mathcal{H}_L = K\mathcal{S}_L \subset \mathcal{H}_{out}$. The out representation Fock space will have a

[1]The importance of Cauchy surfaces will be discussed in section 3.4.

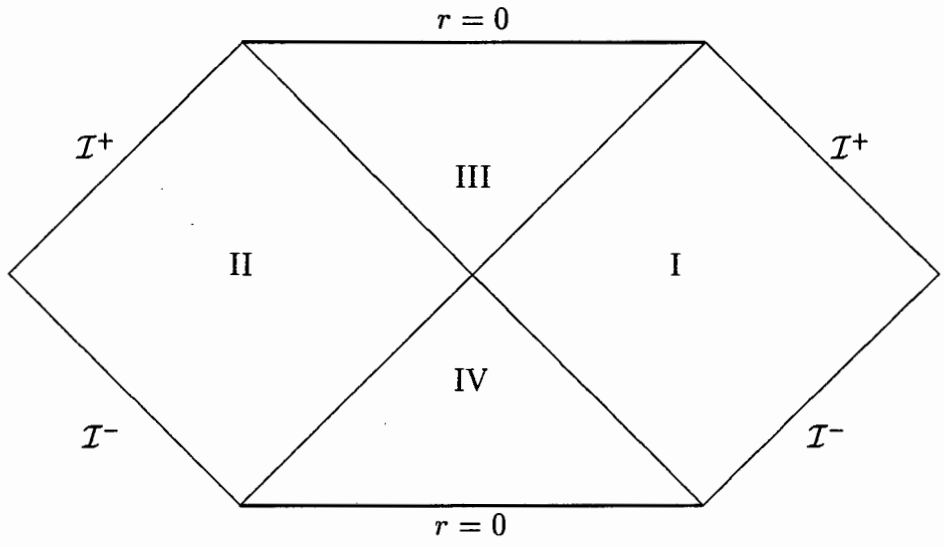

Figure 3.3: A conformal diagram of extended Schwarzschild spacetime showing the singularity at $r = 0$ and the spacelike surface, Σ.

natural particle interpretation for modes in $\mathcal{H}_L$ (with respect to measurements by static observers in the future). This, however, does not hold for $\mathcal{H}_E$, the subspace of $\mathcal{H}_{out}$ associated with $\mathcal{S}_E$.

At this stage, we can compute the S-matrix, $U : \mathcal{F}_s(\mathcal{H}_{in}) \to \mathcal{F}_s(\mathcal{H}_{out})$. We are, as usual, particularly interested in calculating $U|0_{in}>$. The corresponding element in $\mathcal{F}_s(\mathcal{H}_L)$ (note: not $\mathcal{F}_s(\mathcal{H}_{out})$) will inform us about particles seen by static observers at late times.

We start with solutions in $\mathcal{H}_L$ which we then propagate into the past. The solution $\phi \in \mathcal{H}_L$ is positive frequency in the future, but will admit a decomposition into positive and negative frequency parts in the past. It is possible to further decompose $\mathcal{H}_L \subset \mathcal{H}_{out}$ as

$$\mathcal{H}_L = \mathcal{H}_{wh} \oplus \mathcal{H}_{-\infty} \tag{3.6}$$

according to how its elements propagate in the extended Schwarzschild spacetime: $\phi_L \in \mathcal{H}_{wh}$ if $\tilde{\phi}$ propagates entirely through the white hole horizon in the past. On the other hand, $\mathcal{H}_{-\infty}$ consists of solutions, ϕ_L, such that in the

asymptotic past, $\tilde{\phi}_L$ propagates entirely to infinity in region I of extended Schwarzschild spacetime.

The latter is not very interesting for our purposes: the solutions in $\mathcal{H}_{-\infty}$ never enter the non-static region, i.e. they do not interact with the collapsing matter at all (they either get swallowed up by the black hole, or reach $\mathcal{I}^+$, and cannot get their frequencies scrambled, as it were). Thus no particle creation is possible for these modes. Technically, this means that

$$DC^{-1}\phi_{-\infty} = 0 \tag{3.7}$$

Now when we propagate $\phi_{wh} \in \mathcal{H}_{wh}$ backwards into the past, it passes through the collapsing matter (thus suffering interaction with the spacetime geometry), and emerges as an ultra-high frequency solution. [This is almost by definition, since it is only high frequency solutions that can get through by overcoming the large red shift present just before the black hole is formed.] The decomposition of ϕ_{wh} into positive and negative frequency parts (in the aymptotic past) is identical with the decomposition of $\tilde{\phi}_{wh} \in \mathcal{S}^S$ into positive and negative frequency parts with respect to "affine parameter" along h_A. So what we now require is the relationship between the notions of "positive frequency" with respect to affine and Killing times along the generators of a Killing horizon. This will enable us to evaluate functions with respect to both parameters on h_A, which, as already stated, is a Killing horizon.

So consider the bifurcate Killing horizon shown in fig.(3.1) above. Let U, and u, respectively denote the affine and Killing times on H^-, so that $U = e^{-\kappa u}$ on H^- to the causal future of the bifurcation surface Σ and $U = -e^{\kappa u}$ to the causal past of Σ. On H^- consider the following two functions

$$f_{1\omega} \equiv \begin{cases} = g(s)e^{-i\omega u} = g(s)e^{[i\omega\kappa^{-1}ln(-U)]} & U < 0 \\ = 0 & U > 0 \end{cases}$$

$$f_{2\omega} \equiv \begin{cases} = 0 & U < 0 \\ = g(s)e^{i\omega u} = g(s)e^{[-i\omega\kappa^{-1}ln(U)]} & U > 0 \end{cases}$$

where g(s) denotes an arbitrary (smooth) function on Σ, i.e. g is constant along each generator of H^-. Then $f_{1\omega}$ and $f_{2\omega}$ are purely positive frequency functions with respect to Killing time u. (The reader willing to accept this statement is advised to skip over the following argument, and may resume reading the paragraph following eqn 3.12.) Let us see how: The Fourier transform of $f_{1\omega}$ with respect to U is

$$\begin{aligned} \hat{f}_{1\omega} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\sigma U} f_{1\omega}(U) dU \\ &= \frac{1}{\sqrt{2\pi}} g(s) I(\sigma) \end{aligned} \tag{3.8}$$

where

$$I(\sigma) = \int_0^{\infty} e^{i\sigma U} e^{-\frac{i\omega}{a} lnU} dU \tag{3.9}$$

The essential point now is to compare $I(\sigma)$ with $I(-\sigma)$. To do so, we extend the ln function to the complex plane with its branch cut along the negative real axis. For $\sigma > 0$, let $U = iy$, with $y \in I\!R^+$. Since

$$\ln U = \ln(iy) = \frac{i\pi}{2} + \ln y \tag{3.10}$$

it follows that

$$I(\sigma) = ie^{\pi\omega/2a} \int_0^{\infty} e^{-\sigma y} e^{-\frac{i\omega}{a} lny} dy \tag{3.11}$$

Performing a similar analysis for $I(-\sigma)$ we finally obtain that

$$\hat{f}_{1\omega}(-\sigma) = -e^{-\frac{\pi\omega}{a}} \hat{f}_{1\omega}(\sigma) \tag{3.12}$$

So when Fourier analysed with respect to affine parameter U, both $f_{1\omega}$ and $f_{2\omega}$ have nonvanishing negative frequency parts! However, the following linear combinations of $f_{1\omega}$ and $f_{2\omega}$:

$$F_{1\omega} = f_{1\omega} + e^{-\pi\omega/\kappa} f^*_{2\omega} \tag{3.13}$$

$$F_{2\omega} = f_{2\omega} + e^{-\pi\omega/\kappa} f^*_{1\omega} \tag{3.14}$$

are, in fact, positive frequency with respect to U, since $\hat{F}_{1\omega}(-\sigma) = 0$. Now take these functions (together with the condition of no incoming radiation

from infinity) as initial data for solutions, ψ_{wh} and ψ_{II}, in the collpasing body spacetime. We use the solutions to form an orthonormal basis for $\mathcal{H}_{wh}$ and $\mathcal{H}_{II}$, denoted $\{\psi_{iwh}\}$ and $\{\psi_{i\infty}\}$ respectively. Then the solutions

$$\tilde{\Psi}_\omega = \tilde{\psi}_{wh} + e^{-\frac{\pi\omega}{\kappa}}\bar{\tilde{\psi}}_{II} \tag{3.15}$$

and

$$\tilde{\Psi}'_\omega = \tilde{\psi}_{II} + e^{-\frac{\pi\omega}{\kappa}}\bar{\tilde{\psi}}_{wh} \tag{3.16}$$

are purely positive frequency with respect to affine parameter on the white hole horizon. $\tilde{\psi}_{II}$ has support in region II of extended Schwarzschild spacetime, and is obtained from $\bar{\psi}_{wh}$ by using the wedge-reflection isometry present in that spacetime. Then in the spacetime of the collapsing body, the solutions

$$\Psi_\omega = \psi_{wh} + e^{-\frac{\pi\omega}{\kappa}}\bar{\psi}_{II} \tag{3.17}$$

$$\Psi'_\omega = \psi_{II} + e^{-\frac{-\pi\omega}{\kappa}}\bar{\psi}_{wh} \tag{3.18}$$

are purely positive frequency in the asymptotic past. ψ_{II} will, in fact, enter the black hole at "early times". This helps us to define the early time subspace, $\mathcal{H}_E \subset \mathcal{H}_{out}$ as consisting of all ψ_{II} with ψ_{wh} either reaching infinity or entering the black hole at late times. We can write down the S-matrix, U, straightaway, using the techniques of section 2.3. It is obvious that

$$C\Psi_\omega = \psi_{wh} \qquad C\Psi'_\omega = \psi_{II} \tag{3.19}$$

$$D\Psi_\omega = e^{-\frac{\pi\omega}{\kappa}}\bar{\psi}_{II} \quad D\Psi_\omega = e^{-\frac{\pi\omega}{\kappa}}\bar{\psi}_{wh} \tag{3.20}$$

From this we obtain that

$$DC^{-1}\psi_{wh} = e^{-\frac{\pi\omega}{\kappa}}\bar{\psi}_{II} \tag{3.21}$$

$$DC^{-1}\psi_{II} = e^{-\frac{\pi\omega}{\kappa}}\bar{\psi}_{wh} \tag{3.22}$$

We have thus determined the operator, $\mathcal{E} = \bar{D}\bar{C}^{-1}$ and the corresponding two-particle state

$$\varepsilon^{ab} = \prod_i e^{\frac{\pi\omega}{\kappa}} 2(\psi_{wh})^{(a}(\psi_{II})^{b)}$$

The out-state corresponding to the 'in' vacuum, $|0_{in}>$ may thus be written as

$$U|0_{in}>=\prod_i e^{-\frac{n\pi\omega}{\kappa}}|n_{wh}>\otimes|n_{II}>$$

where $|n_{wh}>\in \mathcal{F}_s(\mathcal{H}_{wh})$ denotes an n-particle state corresponding to the mode $|\psi_{wh}>$, and similarly for $|n_{II}>$. By tracing out over the modes $|\psi_{II}>$ which propagate entirely inside the black hole, we obtain the density matrix, ρ, associated with the white hole modes

$$\rho=\prod_i\{\Sigma_i^\infty e^{-\frac{2\pi n\omega_i}{\kappa}}|n_{wh}><n_{wh}|\}$$

We are led to conclude that the state $U|0_{in}>$ is a thermal state at the Hawking temperature $T=\frac{\kappa}{2\pi}$ with respect to the solutions ψ_{wh}. It is a vacuum state with respect to the solutions in $\mathcal{H}_{-\infty}$ We may say that the particle states corresponding to solutions which emerge at late times from the "white hole horizon" in extended Schwarzschild spacetime are thermally populated. At late times, an observer would see a thermal flux of particles appearing to emanate from the black hole.

3.3 Backreaction: Outline of the problem

Our program henceforth is to examine some of the intricacies associated with the semiclassical backreaction program. What the Hawking effect has shown is a definite interaction between the gravitational field and quantum fields propagating in certain spacetimes. This analysis is weighed heavily in favour of the effect the gravitational field has on the quantum ones. We would expect that as a result of the particle creation, the black hole would undergo an energy loss; and energy being a source of curvature, one would naturally expect this loss to effect some change in the black hole. As the name suggests, the aim of studying backreaction effects then is to determine how these quantum fields affect the spacetime, in turn. The appellation 'semiclassical' results from the fact that these backreaction effects are examined within the

framework of the modified Einstein equation:

$$G_{\mu\nu} = 8\pi < \hat{T}_{\mu\nu} > \tag{3.23}$$

where the left hand side is the usual classical Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$, while the right hand side stands for (the expectation value in some still to-be-determined state of) the quantum stress-energy tensor, $\hat{T}_{\mu\nu}$. [The circumflex here highlights the quantum nature of this tensor; we shall drop it henceforth except where confusion may arise, and take it as given that any reference to $T_{\mu\nu}$ means the quantum operator and $< T_{\mu\nu} >$ the expectation value of the quantum operator]. This coupling between a quantum object and a classical one is not without precedent. Before the final formulation of quantum electrodynamics (QED), some success was achieved in treating the elctromagnetic field classically with the quantum fermionic field coupled to it via the analogue of eqn. 3.23 above. As for the range of validity of the semiclassical approach, one would expect it to hold in regions where the curvature is well below the Planck scale. In addition, certain conditions are expected to hold on the stress tensor: the quantum fluctuations are not 'too large'; precisely [42], this means that

$$< T_{\alpha\beta}(x)T_{\mu\nu}(y) > \approx < T_{\alpha\beta}(x) >< T_{\mu\nu}(y) > \tag{3.24}$$

Notwithstanding these statements, one can adopt the viewpoint [18] that not even for Schrödinger quantum mechanics do we have anything more than a 'semiclassical' theory: in interference experiments, one does not take into account the quantum nature of the photographic plates or the main experimental apparatus, it is only the appropriate particle that we analyse quantum mechanically. Similarly, the semiclassical approach to the backreaction program is a more than useful approach to a fuller understanding that will presumably come with a quantum theory of gravity.

We turn now to explicit calculations of the stress-tensor. Now, perhaps not unexpectedly, since we are dealing with quantum fields we are not immune to the infinities which arise in most, if not all, quantum field theories. In our case, these infinities arise in 'naive' attempts to calculate expectation

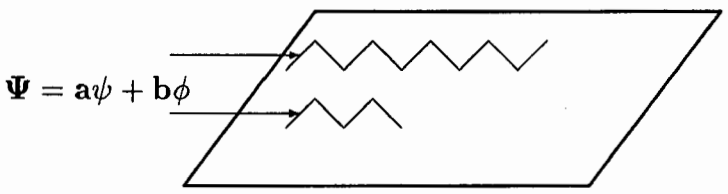

Figure 3.4: Even a wavefunction impinging on a photographic plate is semi-classical!

values of the stress-tensor. By this we mean that a straightforward substitution of classical quantities by their quantum counterparts in the stress-tensor is doomed to failure!

The classical stress-tensor for the Klein-Gordon field, φ is given by

$$T_{\mu\nu} = \nabla_\mu\varphi\nabla_\nu\varphi - \frac{1}{2}g_{\mu\nu}\nabla_\sigma\varphi\nabla^\sigma\varphi \tag{3.25}$$

so that for the quantum operator, $T_{\mu\nu}$, the following, purely formal expression holds:

$$T_{\mu\nu} = \frac{1}{2}(\nabla_\mu\varphi\nabla_\nu\varphi + \nabla_\mu\varphi\nabla_\nu\varphi) - g_{\mu\nu}\nabla_\sigma\varphi\nabla^\sigma\varphi \tag{3.26}$$

where φ is now the quantum field operator. When the gravitational field has the requisite asymptotic properties necessary to define particles, then with the decomposition of the field as

$$\varphi = \Sigma_i(G_i a_i + \bar{G}_i a_i^\dagger) \tag{3.27}$$

we obtain for the stress-tensor the rather long-winded expression

$$\begin{aligned} T_{\mu\nu} = \frac{1}{2}\Sigma_{i,j}\{&[2\nabla_\mu G_i \nabla_\nu G_j - g_{\mu\nu}\nabla_\sigma G_i \nabla^\sigma G_j]a_i a_j \\ &+[\nabla \bar{G}_i \nabla_\nu G_j + \nabla_\nu \bar{G}_i \nabla_\mu G_j - g_{\mu\nu}\nabla_\sigma \bar{G}_i \nabla^\sigma G_j]a_i^\dagger a_j \\ &+[\nabla_\mu G_i \nabla_\nu \bar{G}_j + \nabla_\nu G_i \nabla_\mu \bar{G}_j - g_{\mu\nu}\nabla_\sigma G_i \nabla^\sigma \bar{G}_j]a_i a_j^\dagger \\ &+[2\nabla_\mu \bar{G}_i \nabla_\nu G_j - g_{\mu\nu}\nabla_\sigma \bar{G}_i \nabla^\sigma \bar{G}_j]a_i^\dagger a_j^\dagger\} \end{aligned} \tag{3.28}$$

If we wanted to calculate the expectation value of this quantity in say, some vacuum state $|0_{in}>$, then

$$< 0_{in}|T_{\mu\nu}|0_{in} >= \frac{1}{2}\Sigma_i[\nabla_\mu G_i \nabla_\nu \bar{G}_i + \nabla_\nu G_i \nabla_\mu \bar{G}_i - g_{\mu\nu}\nabla_\sigma G_i \nabla^\sigma \bar{G}_i] \tag{3.29}$$

This sum does not converge, as was the case with the field in eqn.(3.27) anyway, as we saw earlier. The problem is compounded by the fact that, since (in our rigourous reformulation) φ is well defined only as a distribution on spacetime, then eqn. 3.26 above involves taking the product of two distributions at the same spacetime point, a totally nebulous procedure mathematically. [However, for orthogonal states such as the vacuum state $|0_{in}\rangle$ and a 2-particle in-state $|kl\rangle$, we have that

$$\langle kl|T_{\mu\nu}|0_{in}\rangle = \frac{1}{2}\{\nabla_\mu \bar{G}_k \nabla_\nu \bar{G}_l + \nabla_\nu \bar{G}_k \nabla_\mu \bar{G}_l - g_{\mu\nu}\nabla_\sigma \bar{G}_k \nabla^\sigma \bar{G}_l\} \tag{3.30}$$

which is manifestly *sans* summation, and thus finite.] The way to deal with these infinities is with the technique of renormalisation, to which we now turn our attention.

3.3.1 The Effective Action.

The classical Einstein equation may be derived from the action

$$S = S_g + S_m \tag{3.31}$$

using the principle

$$\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} = 0. \tag{3.32}$$

The gravitational part of the action in eqn.(3.31) is given by

$$S_g = \int (\frac{\sqrt{-g}}{16\pi G_B}) R d^4x \tag{3.33}$$

from which we obtain (using $\frac{2}{\sqrt{-g}} \frac{\delta S_g}{\delta g^{\mu\nu}} = 0$) the left hand side of the classical Einstein equation:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = -8\pi G T_{\mu\nu}. \tag{3.34}$$

The right hand side of this equation is obtained by varying the S_m term in 3.31 above. In the semiclassical picture we would like to work in, we have to replace the right hand side of 3.34 by the expectation value of the stress energy tensor, so that

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = -8\pi G_B < T_{\mu\nu} > . \tag{3.35}$$

To proceed as in the classical case using the ideas of the action, we seek an object, W, called the *effective action* for the quantum matter fields, which, when functionally differentiated, yields the desired $< T_{\mu\nu} >$, so that:

$$\frac{2}{\sqrt{-g}} \frac{\delta W}{\delta g^{\mu\nu}} =< T_{\mu\nu} > . \tag{3.36}$$

Using the techniques of path integral quantization, we may obtain W as

$$W = -\frac{1}{2} i \text{ tr}[ln(-G_F)] \tag{3.37}$$

where G_F is the Feynman Green function, which is interpreted as an operator which acts on a space of vectors $|x >$, in such a way that

$$G_F(x, x') =< x|G_F|x' > . \tag{3.38}$$

It turns out that we must use a representation for G_F. We use the DeWitt-Schwinger representation, denoted G_F^{DS}, which is defined using an adiabatic approximation [8] as

$$G_F^{DS}(x, x') = -\Delta^{\frac{1}{2}}(x, x')(4\pi)^{-n/2} \int_0^\infty ids(is)^{-n/2} e^{-im^2 s + (\sigma/2is)} F(x, x'; is) \tag{3.39}$$

Δ is the Van Vleck determinant, σ here indicates the geodesic distance between x and x' and $F(x,x';is) \approx a_0(x,x') + a_1(x,x')(is)^2 + \cdots$. The nature of the coefficients a_0, $a_1 \cdots$ will be revealed shortly. This formula is exact (despite the asymptotic appearance of F). However, when we extend the expansion of F to all (adiabatic) orders as

$$F(x,x';is) \approx \Sigma_{j=0}^{\infty} a_j(x,x')(is)^j \tag{3.40}$$

with $a_0 = 1$, then we may write

$$G_F^{DS} \approx \frac{-i\pi\Delta^{\frac{1}{2}}(x,x')}{4\pi i}^{n/2} \Sigma_{j=0}^{\infty} a_j(x,x')(-\frac{\partial}{\partial m^2})^j \times[\frac{2m^{2\,(n-2)/4}}{-\sigma} H^{(2)}_{(n-2)/2}((2m^2\sigma)^{\frac{1}{2}})] \tag{3.41}$$

So

$$< x|ln(-G_F^{DS})|x' >= -\int_{m^2}^{\infty} G_F^{DS}(x,x')dm^2 \tag{3.42}$$

from which follows

$$W = \frac{1}{2}i\int_{m^2}^{\infty} dm^2 \int d^n x\sqrt{-g(x)}G_F^{DS}(x,x) \tag{3.43}$$

where the limit $x' \to x$ has been taken. Now W being an action, known as the *one-loop effective action* (since it is evaluated with no sources present), we may define an effective Lagrangian

$$L_{eff}(x) = \frac{1}{2}i \lim_{x' \to x} \int_{m^2}^{\infty} dm^2 G_F^{DS}(x,x'). \tag{3.44}$$

L_{eff} contains certain divergences related to the representation we are using. Specifically, when the limit $x' \to x$ is taken, the 'damping factor' $\sigma/2s$ in the exponent of G_F^{DS} vanishes, leaving a divergence at the lower end of the integral. We may identify the (potentially) divergent terms as

$$L_{div} = -\lim_{x' \to x} \frac{\Delta^{\frac{1}{2}}(x,x')}{32\pi^4} \int_0^{\infty} \frac{ds}{s^3} e^{-i(m^2 s - \sigma/2s)} \times[a_0(x,x') + a_1(x,x')is + a_2(x,x')(is)^2] \tag{3.45}$$

These divergences in $L_{eff}(x)$ will naturally be reflected in $< T_{\mu\nu} >$. Physically, the divergences arise from the short wavelength behaviour of the field, i.e. the ultrahigh frequency modes. Thus these divergences are a probe of the local structure of the spacetime. But, being local and thus geometric in nature, these terms L_{div} are more correctly seen as gravitational terms! In this observation lies the key to the eventual renormalisation of the stress-tensor. What we would like to achieve is the comparison of terms in L_{div} with certain terms in the classical gravitational Lagrangian, L_g. To do so, we have to *regularise* the expressions, with the eventual aim of absorbing them into L_g.

One way to do this is to examine the effective Lagrangian for divergences:

$$L_{eff} \approx \lim_{x' \to x} \frac{(\Delta)^{\frac{1}{2}}(x,x')}{2(4\pi)^{n/2}} \Sigma_{j=0}^{\infty} a_j(x,x') \int_0^{\infty} (is)^{j-1-n} e^{-i(m^2 s - \frac{\sigma}{2s})} ids \qquad (3.46)$$

of which the first $\frac{1}{2}n + 1$ terms are divergent as $\sigma \to 0$. If we analytically continue n to complex values, then as $x' \to x$, we obtain

$$L_{eff} = \frac{1}{2}(4\pi)^{-n/2} \Sigma_{j=0}^{\infty} a_j(x)(m^2)^{n/2-j} \Gamma(j - n/2) \qquad (3.47)$$

This makes the divergent terms manifest: as $n \to 4$, the first 3 terms blow up because of pole in the Γ functions, e.g. $j = 0 : \Gamma(-n/2) = \frac{4}{n(n-2)}(\frac{2}{4-n} - \gamma) + O(n-4)$, where γ is Euler's constant. Then we write for these three divergent terms

$$L_{div} = -(4\pi)^{-\frac{n}{2}} \{ \frac{1}{n-4} + \frac{1}{2}[(\gamma + ln(\frac{m^2}{\mu^2})] \} \qquad (3.48)$$

where μ is an arbitrary scale.

That L_{div} is an entirely geometric object may be seen by examining the functions $a_i(x)$; for example

$$a_2(x) = \frac{1}{180} R_{\alpha\beta\gamma\delta} R^{\alpha\beta\gamma\delta} - \frac{1}{180} R^{\alpha\beta} R_{\alpha\beta} - \frac{1}{6}(\frac{1}{5} - \zeta)\Box R + \frac{1}{2}(\frac{1}{6} - \zeta)^2 R^2 \qquad (3.49)$$

As part of the geometry, we may try to absorb L_{div} into the rest of the total Lagrangian, i.e. into L_{grav}. Recall that the action for the latter is given by

$$S_g = \int \sqrt{-g} \frac{1}{16\pi G_B} (R - 2\Lambda) d^n x \qquad (3.50)$$

so that the modified Lagrangian density becomes

$$L_g^{mod} = \sqrt{-g}[-(A+\frac{\Lambda_B}{8\pi G_B})+(B+\frac{1}{16\pi G_B})R-\frac{a_2(x)}{(4\pi)^{n/2}}\{\frac{1}{n-4}+\frac{1}{2}[\gamma+ln(\frac{m^2}{\mu^2})]\}} \tag{3.51}$$

The precise values of A and B are not of interest now; what does attract our attention is the fact that the first term in L_g^{mod} above is a constant, i.e. A and $\frac{\Lambda_B}{8\pi G_B}$ are part of a more general object, $\Lambda = \Lambda_B + 8\pi G_B A$. Thus we say that the scalar field (from which the matter Lagrangian is constructed in the first place), 'dresses' the 'bare' cosmological constant, Λ_B. Similar remarks apply to the Ricci scalar and B, which give rise to a renormalised gravitational constant, $G = \frac{G_B}{1+16\pi G_B B}$.

Now the last term on the right hand side is new (i.e. not part of the classical Einstein equation). With this, the left hand side of the field equation becomes

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} + \alpha^{(1)}H_{m\nu} + \beta^{(2)}H_{\mu\nu} + \gamma H_{\mu\nu} \tag{3.52}$$

(the superscripts, (1) and (2) belonging to the $H_{\mu\nu}$). This might seem rather intimidating until one realises that the $H_{\mu\nu}$, ${}^{(1)}H_{\mu\nu}$ and ${}^{(2)}H_{\mu\nu}$ terms are constructed from geometrical scalars, such a scalar being an invariant formed from the curvature scalar and its contractions (just like the Ricci scalar); e.g.

$${}^{(2)}H_{\mu\nu} \equiv \frac{1}{\sqrt{-g}}\frac{\delta}{\delta g^{\mu\nu}}\int \sqrt{-g}R^{\alpha\beta}R_{\alpha\beta}d^n x \tag{3.53}$$

So what we have done is to render the divergent L_{div} terms in form which appears finite; this is called regularisation. [The technique used above of allowing n to become complex is called dimensional regularisation]. The last step is to obtain the renormalised (finite) effective Lagrangian, L_{ren}, describing the matter:

$$L_{ren} = L_{eff} - L_{div} \tag{3.54}$$

It turns out that

$$L_{ren} = \frac{1}{64\pi^2}\int_0^\infty ln(is)\frac{\partial^3}{\partial(is)^3}[F(x,x';is)e^{-ism^2}]ids \tag{3.55}$$

There exist other techniques, such as ζ-function regularisation[26], to obtain a renormalised L_{ren}. We shall not discuss these. To conclude this section, we summarise: when we look at the total semiclassical gravitational- matter system, the action is of the form $S = S_g + W$. When divergences in W become apparent, they are absorbed into S_g, thus modifying the above to $S = (S_g)_{ren} + W_{ren}$ where W_{ren} is now finite.

Remarks

- Despite the elegance of this method showing the divergences in the effective action, it is almost impossible to carry out in practice to obtain the requisite stress tensor. This is due to the functional differentiation required to obtain $< T_{\mu\nu} >_{ren}$ from W_{ren}, as in

$$\frac{2}{\sqrt{-g}}\frac{\delta W}{\delta g_{\mu\nu}} =< T_{\mu\nu} > \tag{3.56}$$

  We shall thus turn our attention in the following section to the more practical approach of obtaining a suitable $< T_{\mu\nu} >$ by the technique of point splitting.

- The classical Einstein equation is of second order in derivatives of the metric. However, the method employed here of obtaining $< T_{\mu\nu} >$ made use of an (adiabatic) approximation which required terms of fourth order in derivatives of the metric. This leads to unacceptable, 'runaway' solutions[54].

3.4 Backreaction II: The algebraic approach

L'etat c'est moi. King Louis XIV of France

To investigate the renormalisation of the stress tensor in greater depth, we focus attention on the algebraic approach to quantum field theory. It may seem that by doing so, we have abandoned our intention to examine physics from a geometric point of view. Technically, this may be correct, but the

spirit of the algebraic approach to quantum field theory is similar to the ideas of (coordinate free) differential geometry. More precisely, we free the notion of 'state' from the confines of its Fock space interpretation. The crucial difference results from the reversal of the roles played by observables and states. In the Fock space approach, observables are operators which "act upon" the state. Using the algebraic approach, one constructs observables as elements of an abstract algebra. States then act upon observables (in other words, they are dual to the latter); this action is similar to taking expectation values in the Fock space treatment. The advantage gained is that we are able to deal quite effectively with the unitarily inequivalent quantum field theories, since with each of them we may associate a certain state. But states (all states) are exactly what we deal with in the algebraic approach, and thus we are able to defeat the deficiencies of the Fock space approach. The rigour afforded by the algebraic approach also allows the formulation of such crucial issues as the formulation of quantum field theory in spacetimes with closed timelike curves [16].

To proceed, we first remark that the key element to note is that the algebraic structure of the field operators are "identical" for unitarily inequivalent constructions (which is exactly what we want, viz that such constructions be characterised by *states*, as noted above). So what we have to do is establish the necessary algebra and then proceed to construct the states.

3.4.1 The Klein-Gordon field and the notion of state

We shall henceforth concern ourselves exclusively with the quantum theory of the linear Klein-Gordon field in globally hyperbolic spacetimes. These spacetimes are distinguished by the fact that they contain Cauchy surfaces. This permits the introduction of a useful notion of 'time'. (Interest in the non-globally hyperbolic case relate also to spacetimes with closed timelike curves [16]). As for the focus on linear fields, the results we are interested in are difficult to extend to the non-linear case. We will focus in particular on the case of a real scalar field, $\varphi : \mathcal{M} \to I\!R$. But it is worth remarking that this analysis is extendible to general bosonic fields, as well as to linear, real, fermionic fields with the requisite modifications (mainly related to the

question of signs). It will be important to note for what we are about to do that the Klein-Gordon equation

$$(\Box_g + \mu^2)\Phi = 0 \tag{3.57}$$

is a hyperbolic differential equation for which the Cauchy problem on a globally hyperbolic spacetime is well posed. (We rely heavily here in parts on the analysis of[30].) Thus there exist [13] continuous linear operators

$$\Delta_{R,A} : \Delta\mathcal{M} \to \mathcal{E}(\mathcal{M})$$

with the properties

$$\begin{aligned} &(\Box_g + \mu^2)\Delta_{R,A} f = \Delta_{R,A}(\Box_g + \mu^2) f = f \\ &supp\,(\Delta_A f) \subset J^-(supp\, f) \\ &supp\,(\Delta_R f) \subset J^+(supp\, f) \end{aligned} \tag{3.58}$$

for $f \in \mathcal{D}(\mathcal{M})$. These are the familiar (see eqns. 2.53, 2.54 above) advanced (Δ_A) and retarded (Δ_R) fundamental solutions of the Klein-Gordon equation (3.57), while $E := \Delta_R - \Delta_A$ is called the fundamental solution or propagator of (3.57). (For general non-globally hyperbolic spacetimes, there is no advanced-minus-retarded fundamental solution [16].) It has the properties

$$\begin{aligned} &(\Box_g + \mu^2)Ef = E(\Box_g + \mu^2) f = 0 \\ &supp\,(Ef) \subset J^+(supp\, f) \cup J^-(supp\, f) \end{aligned} \tag{3.59}$$

for $f \in \mathcal{D}(\mathcal{M})$. Δ_R, Δ_A and E can be continuously extended to the adjoint operators

$$\Delta_R', \Delta_A', E' : \mathcal{E}'(\mathcal{M}) \to \mathcal{D}'(\mathcal{M})$$

by $\Delta_R' = \Delta_A$, $\Delta_A' = \Delta_R$, $E' = -E$. Let Σ denote a given Cauchy surface of $\mathcal{M}$ with future-directed unit-normalfield n^α. Then there are the restriction operators

$$\begin{aligned} \rho_o : \quad \mathcal{E}(\mathcal{M}) &\to \mathcal{E}(\Sigma) \\ f &\mapsto f|_\Sigma \\ \rho_1 : \quad \mathcal{E}(\mathcal{M}) &\to \mathcal{E}(\Sigma) \\ f &\mapsto (n^\alpha \nabla_\alpha f)|_\Sigma, \end{aligned} \tag{3.60}$$

which have adjoints ρ_o', ρ_1' mapping $\mathcal{E}'(\Sigma)$ to $\mathcal{E}'(\mathcal{M})$. Dimock [13] proved the following existence and uniqueness result for the Cauchy problem:

Theorem 3.2 *a) $E\rho_o', E\rho_1'$ restrict to continuous operators from $\mathcal{D}(\Sigma)$ ($\subset \mathcal{E}'(\Sigma)$) to $\mathcal{E}(\mathcal{M})$ ($\subset \mathcal{D}'(\mathcal{M})$) and the unique solution of the Cauchy problem (3.57) with initial data $u_o, u_1 \in \mathcal{D}(\Sigma)$ is given by*

$$u = E\rho_o' u_1 - E\rho_1' u_o. \tag{3.61}$$

b) Furthermore, (3.61) also holds in the sense of distributions, i.e. given $u_o, u_1 \in \mathcal{D}'(\Sigma)$, there exists a unique distribution $u \in \mathcal{D}'(\mathcal{M})$ which is a (weak) solution of (3.57) and has initial data $u_o = \rho_o u, u_1 = \rho_1 u$. It is given by

$$u(f) = -u_1(\rho_o E f) + u_o(\rho_1 E f) \tag{3.62}$$

for $f \in \mathcal{D}(\mathcal{M})$.

Applying ρ_o and ρ_1 to the identity (3.61) we immediately obtain:

$$\begin{array}{ll} \rho_o E\rho_0' = 0 & \rho_o E\rho_1' = -1 \\ \rho_1 E\rho_o' = 1 & \rho_1 E\rho_1' = 0 \end{array} \tag{3.63}$$

Inserting $u = Ef$ into both sides of equ. (3.61) we get the identity

$$E = E\rho_o'\rho_1 E - E\rho_1'\rho_o E. \tag{3.64}$$

Theorem 3.2 allows us to formulate the classical phase space of the field theory in terms of initial data on a Cauchy surface. Let Σ be a Cauchy surface for $(\mathcal{M}, g)$ with volume element $d^3\sigma$. Then we define the classical phase space of the Klein-Gordon field as the real linear symplectic space (Γ, σ), where $\Gamma := \mathcal{D}(\Sigma) \oplus \mathcal{D}(\Sigma)$ is the space of initial data with compact support and σ is the symplectic bilinear form

$$\begin{array}{rcl} \sigma : \Gamma \times \Gamma & \to & I\!R \\ (F_1, F_2) & \mapsto & -\int_\Sigma [u_1 p_2 - u_2 p_1]\, d^3\sigma \end{array} \tag{3.65}$$

for $F_i := \binom{u_i}{p_i} \in \Gamma, i = 1, 2$.
(3.65) is independent of the choice of Cauchy surface: If Σ_1 and Σ_2 are two

Cauchy surfaces (enclosing the volume $V \subset \mathcal{M}$) and Φ_1, Φ_2 the solutions of (3.57) to the initial data F_1, F_2 on Σ_1 (with compact supports) then we can write (3.65) as

$$\begin{aligned} -\sigma(\Phi_1, \Phi_2) &= \int_{\Sigma_1} [\Phi_1 \nabla_\alpha \Phi_2 - \Phi_2 \nabla_\alpha \Phi_1] n^\alpha \, d^3\sigma \\ &\equiv \int_{\Sigma_1} j_\alpha n^\alpha \, d^3\sigma \\ &= \int_V (\nabla^\alpha j_\alpha) \, d^4\mu + \int_{\Sigma_2} j_\alpha n^\alpha \, d^3\sigma \\ &= \int_{\Sigma_2} j_\alpha n^\alpha \, d^3\sigma \\ &\equiv \int_{\Sigma_2} [\Phi_1 \nabla_\alpha \Phi_2 - \Phi_2 \nabla_\alpha \Phi_1] n^\alpha \, d^3\sigma \end{aligned}$$

where j_α is the conserved current ($\nabla^\alpha j_\alpha = 0$) of the Klein-Gordon field.
We construct the algebra of the observables as follows: we have in hand both the smeared field operator $\sigma(\phi, \cdot)$ and the identification of the space of solutions $\mathcal{S}$ with the phase space, Γ. An algebra of observables is the $*-$algebra, $\mathcal{A}'$ over $\mathbb{C}$ generated by the objects $\sigma(\hat{\varphi}, \phi)$ and identity element 1 satisfying

$$\begin{aligned} \sigma(\hat{\varphi}, \phi)^* &= \sigma(\hat{\varphi}, \phi) \\ \sigma(\hat{\varphi}, \phi_1), \sigma(\hat{\varphi}, \phi_2) &= i\sigma(\phi_1, \phi_2) \end{aligned} \tag{3.66}$$

$\mathcal{A}'$ consists of equivalence classes of all finite (complex) linear combinations of finite products of the objects $\sigma(\hat{\varphi}, \phi)$. If $\phi \in \mathcal{S}$ has initial data (f, p) on a Cauchy surface Σ, we interpret $\sigma(\hat{\varphi}, \phi)$ as corresponding to

$$\sigma(\hat{\varphi}, \phi) = \hat{\varphi}(p) - \hat{\pi}(f) \tag{3.67}$$

with $\hat{\varphi}(p)$ and $\hat{\pi}(f)$ interpreted as "three-smeared operators" on Σ, representing the field and its momentum. We can write

$$\begin{aligned} \hat{\varphi}(p) &= \int_\Sigma \hat{\varphi}(x) p(x) \sqrt{h} d^3x \\ \hat{\pi}(f) &= \int_\Sigma \hat{\pi} f(x) \sqrt{h} d^3x \end{aligned} \tag{3.68}$$

Eqns. 3.66 become the "equal time" commutation relations

$$[\hat{\varphi}(x), \hat{\pi}(y)] = i\delta(x, y) \tag{3.69}$$

$\mathcal{A}'$ is a $*-$algebra; we would prefer to work with a C^*-algebra. This is achieved by defining the Weyl operators

$$W(\phi) = e^{-i\sigma(\hat{\varphi},\phi)} \tag{3.70}$$

i.e. exponentiated field operators (and thus unitary, see section 1.1), from which follow the Weyl relations

$$\begin{aligned} W(\phi)^* &= W(-\phi) \\ W(\phi_1, \phi_2) &= e^{-i\sigma(\phi_1,\phi_2)/2} W(\phi_1 + \phi_2) \end{aligned} \tag{3.71}$$

(analagous to 3.66 above).

We define the Weyl algebra by starting with the formal finite sums $\Sigma_i a_i W(\phi_i)$ with the $*$ operation determined by 3.71 above. This algebra admits a unique C^*- norm and its completion is the Weyl algebra. This is our 'minimal' algebra for dealing with quantum field theory in curved spacetimes. One of the main advantages of working with the Weyl operators is that we circumvent any problems related to the unboundedness of operators [29]: unbounded operators cannot be defined on the entire Hilbert space, but only on a dense subspace thereof; this ultimately results in (infinitely many) inequivalent representations of the commutation relations.

For any test function $F \in \mathcal{D}(\mathbb{R}^4)$, we define the "exponentiated four-smeared" quantum field $e^{-i\hat{\varphi}(F)}$ by

$$e^{-i\hat{\varphi}(F)} = W(EF) \tag{3.72}$$

It follows that

$$\begin{aligned} e^{-i\hat{\varphi}(F_1)} e^{-i\hat{\varphi}(F_2)} &= e^{-i\sigma(EF_1, EF_2)/2} e^{-i\sigma(F_1+F_2} \\ &= e^{-\frac{1}{2}i\int_{\mathcal{M}} F_1 E F_2 \sqrt{-g} d^4x} e^{-i\hat{\varphi}(F_1+F_2)} \end{aligned} \tag{3.73}$$

The C^*-algebra generated by the four-smeared quantities satisfying eqn. 3.73 is isomorphic via eqn. 3.72 to the C^*-algebra of the three-smeared quantities $W(\phi)$ satisfying eqns. 3.71, 3.71; so the four-smeared and three-smeared

notions are equivalent. It will turn out to be more convenient for later work to deal with the four-smeared quantities.

So to the symplectic space (Γ, σ) there is associated (uniquely up to unitary equivalence) a Weyl algebra $\mathcal{A}[\Gamma, \sigma]$ A local algebra $\mathcal{A}(O)$ ($\mathcal{O}$ an open bounded subset of $\mathcal{M}$) is the C^*-algebra generated by the elements $W(\rho_o Ef, \rho_1 Ef)$ with *supp* $f \subset \mathcal{O}$. It is the algebra of quantum observables measurable in the spacetime region $\mathcal{O}$. Then $\mathcal{A}[\Gamma, \sigma] = \overline{\bigcup_{\mathcal{O}} \mathcal{A}(\mathcal{O})}^{C^*}$.

A state ω on a C^*-algebra $\mathcal{A}$ is a positive normalised linear form on $\mathcal{A}$, i.e. a mapping from $\mathcal{A}$ into $\mathbb{C}$ such that
(i) $\omega(\lambda_1 A_1 + \lambda_2 A_2) = \lambda_1 \omega(A_1) + \lambda_2 \omega(A_2)$;
(ii) $\omega(A^*) = \overline{\omega(A)}$;
(iii)$\omega(A) \geq 0$ for $A \in \mathcal{A}^+$ where $\mathcal{A}^+ = \{K \in \mathcal{A} | K = AA^*\}$ (positivity);
(iv) $\omega(I) = 1$ (normalisation)

The set of all states is denoted $\mathcal{T}(\mathcal{A})$. The relationship between this proposal and the Fock space notion of state is simple: given any density matrix $\varrho : \mathcal{F} \to \mathcal{F}$ on a Hilbert space $\mathcal{F}$ which carries a representation $\pi : \mathcal{A} \to \mathcal{L}(\mathcal{F})$ of $\mathcal{A}$ (see appendix B), an algebraic state $\omega : \mathcal{A} \to \mathbb{C}$ is obtained as

$$\omega(A) = tr(\varrho \pi(A)) \tag{3.74}$$

Thus, all states in all quantum field theory constructions that we have outlined in section 2.2 give rise to algebraic states. The converse of this result is the content of our next theorem, named for Gelfand, Neumark and Segal. (The proof is quite instructive as to the construction of Hilbert spaces here, and is thus included.)

Theorem 3.3 *(GNS construction) For each state ω on a C^*- algebra, there exists a representation π_ω of $\mathcal{A}$ in a Hilbert space $\mathcal{H}_\omega$ with cyclic vector Ω_ω such that*

$$\omega(A) = < \Omega_\omega, \pi_\omega(A)\Omega_\omega > \qquad \forall A \in \mathcal{A} \tag{3.75}$$

This representation is denoted by the triple $(\mathcal{H}_\omega, \pi_\omega, \Omega_\omega)$, and is unique up to unitary equivalence.

Proof: As emphasised several times before, the basic requirement for constructing a Hilbert space is a suitable inner product. The trick here is to use the state ω to define an inner product on $\mathcal{A}$ by the relation:

$$< A, B >_{\mathcal{A}} \equiv \omega(A^* B) \tag{3.76}$$

We complete $\mathcal{A}$ in the norm defined by the above inner product to obtain the Hilbert space $\mathcal{H}_\omega$! To complete the picture, the representation $\pi_\omega : \mathcal{A} \rightarrow \mathcal{L}(\mathcal{H}_\omega)$ is obtained by allowing $\mathcal{A}$ to act upon itself by multiplication, which is extended to the newly-constructed $\mathcal{H}_\omega$. Then the desired cyclic vector Φ_ω is just the identity element of $\mathcal{A}$.■

We can recover the usual notions of eigenvectors and eigenvalues via the following definition: an operator B_q is called a (right) eigenvector of (observable) Q with (right) eigenvalue q if

$$\begin{aligned} \omega(B_q^* B_q) &\neq 0 \\ \text{and } \omega(AQB_q) &= q\omega(AB_q) \end{aligned}$$

Also the notion of 'pure' and 'mixed'states may be introduced in the algebraic framework as follows: an algebraic state, ω is said to be *mixed* if it can be expressed as

$$\omega = c_1\omega_1 + c_2\omega_2 \tag{3.77}$$

where $c_1, c_2 > 0$ and $\omega_1 \neq \omega_2$; if ω cannot be so expressed, then it is said to be *pure*. In the GNS construction shown above, the distinction between the two is contained in the statement that the representation π_ω of $\mathcal{A}$ will be irreducible if and only if ω is pure.

We will need the following concept: a folium, Φ, of states on $\mathcal{A}$ is a subset of $\mathcal{T}(\mathcal{A})$ such that

- if $\omega_1, \omega_2 \in \Phi$ and $\lambda_1, \lambda_2 \in I\!R^+$ with $\lambda_1 + \lambda_2 = 1$, then $\lambda_1\omega_1 + \lambda_2\omega_2 \in \Phi$;
- if $\omega \in \Phi$ and $A \in \mathcal{A}$, then $\omega_A(B) \in \Phi$, where $\omega_A(B) = \frac{\omega(A^*BA)}{\omega(A^*A)}$
- Φ is closed with respect to norm convergence in $\mathcal{T}(\mathcal{A})$.

The set $\mathcal{F}(\omega)$ of states ω_p that corresponds to the density matrices ϱ in $\mathcal{H}_\omega$ (from the GNS construction) according to 3.74 is a folium. Before embarking on our main program, we make some comments concerning this notion of state.

Remarks

1. The two point-function of the quantum field in state ω may be defined by

$$< \hat{\sigma}(\phi_1, \cdot), \hat{\sigma}(\phi_2, \cdot) >_\omega = -\frac{\partial^2}{\partial s \partial t}\{\omega[W(s\phi_1 + t\phi_2)e^{ist\sigma(\phi_1\phi_2)}]\}_{s=t=0} \tag{3.78}$$

   Using the correspondence between solutions and test functions given in the Lemma 2.2, we may view the two-point function as a bi-distribution on spacetime, denoted by $< \varphi(x_1)\varphi(x_2) >$. The higher $n-$ point functions are defined similarly. The specification of a state, ω, on $\mathcal{A}$ corresponds roughly to the specification of all the smeared $n-$ point functions of the quantum field. [This is reminiscent of the work discussed in a somewhat different setting, viz section 1.3, where the ray to which the wavefunction, ϕ, belongs is determined by examining the expectation values of 'enough' observables.]

2. Although $\mathcal{A}$ is the 'fundamental' algebra of observables, it is by no means all-embracing; in particular, the stress-tensor of the quantum field is not an element of $\mathcal{A}$. Instead, $\mathcal{A}$ contains only enough observables to formulate the theory. To accomodate additional observables, we would either have to enlarge $\mathcal{A}$ and/or restrict the notion of state.

Before attempting to do so, we state a result which sheds some light on this discussion:

Theorem 3.4 *(Fell): Let $(\mathcal{F}_1, \pi_1)$ and $(\mathcal{F}_2, \pi_2)$ be (possibly unitarily inequivalent) representations of the Weyl algebra $\mathcal{A}$. Let $A_1, \ldots, A_n \in \mathcal{A}$ and let $\epsilon_1 \ldots, \epsilon_n > 0$. Let ω_1 be an algebraic state corresponding to a density matrix*

on $\mathcal{F}_1$. Then there exists a state ω_2 corresponding to a density matrix on $\mathcal{F}_2$ such that for all $i = 1, \ldots, n$ we have

$$|\omega_1(A_i) - \omega_2(A_i)| < \epsilon_i \tag{3.79}$$

This result shows that even for unitarily inequivalent constructions, the collection of states are "physically equivalent" with regard to observables in $\mathcal{A}$ in the sense that the measurements of a finite number of expectation values of observables cannot distinguish between the representations. So it might seem that our work of looking at unitarily inequivalent schemes is without value as far as 'physics' is concerned. However, we are spared this embarassment by the fact that additional observables, such as the stress-tensor, exists outside the domain of $\mathcal{A}$ and thus two representations need not be physically equivalent with respect to them.

We now turn our attention to a restricted class of states in the hope of ultimately obtaining a reasonable (i.e. finite) notion of the stress-tensor.

This is accomplished by restricting the allowable class of states, as prescribed above. For linear systems, we restrict ourselves to the quasifree states, all of whose truncated n-point functions vanish for $n \neq 2$:

Definition 3.5 *Let $\mu : \Gamma \times \Gamma \to \mathbf{R}$ be a real scalar product satisfying*

$$\frac{1}{4}|\sigma(F_1, F_2)|^2 \leq \mu(F_1, F_1)\mu(F_2, F_2) \tag{3.80}$$

for all $F_1, F_2 \in \Gamma$. Then the **quasifree state** *ω_μ associated with μ is given by*

$$\omega_\mu(W(F)) = e^{-\frac{1}{2}\mu(F,F)}. \tag{3.81}$$

If ω_μ is pure it is called a **Fock state.**

A quasifree state with vanishing one-point function is called a Gaussian state. The connection between this algebraic notion of a quasifree state and the usual notion of "vacuum state" in a Hilbert space is established by the following theorem (see [31]):

Theorem 3.6 *Let ω_μ be a quasifree state on $\mathcal{A}[-, \sigma]$.*

a) Then there exists a **one-particle Hilbert space structure**, *i.e. a Hilbert space $\mathcal{H}$ and a real-linear map $K : \Gamma \to \mathcal{H}$ such that*
i) $K\Gamma + iK\Gamma$ is dense in $\mathcal{H}$,
ii) $\mu(F_1, F_2) = \mathrm{Re}\langle KF_1, KF_2\rangle_{\mathcal{H}}$ $\forall F_1, F_2 \in \Gamma$,
iii) $\sigma(F_1, F_2) = 2\mathrm{Im}\langle KF_1, KF_2\rangle_{\mathcal{H}}$ $\forall F_1, F_2 \in \Gamma$.
Moreover, the pair $(K, \mathcal{H})$ is uniquely determined up to unitary equivalence.
It follows that ω_μ is pure. $\iff$ $K(\Gamma)$ is dense in $\mathcal{H}$.

b) The GNS-triple $(\mathcal{H}_{\omega_\mu}, \pi_{\omega_\mu}, \Omega_{\omega_\mu})$ of the state ω_μ can be represented as $(\mathcal{F}^s(\mathcal{H}), \rho_\mu, \Omega^{\mathcal{F}})$, where
i) $\mathcal{F}^s(\mathcal{H})$ is the symmetric Fock space over the one-particle Hilbert space $\mathcal{H}$,
ii) $\rho_\mu[W(F)] = \exp\{-i\overline{[a^*(kF) + a(kF)]}\}$, where a^* and a are the standard creation and annihilation operators on $\mathcal{F}^s(\mathcal{H})$ satisfying

$$[a(u), a^*(v)] = \langle u, v\rangle_{\mathcal{H}} \text{ and } a(u)\Omega^{\mathcal{F}} = 0$$

for $u, v \in \mathcal{H}$ (the bar denotes the closure of the operator).
iii) $\Omega^{\mathcal{F}} := \mathbf{1} \oplus \mathbf{0} \oplus \mathbf{0} \oplus \ldots$ is the (cyclic) Fock vacuum.
It is known that: ω_μ is pure $\Leftrightarrow$ ρ_μ is irreducible.

Thus, ω_μ can also be represented as $\omega_\mu(W(F)) = \exp\{-\frac{1}{2}||kF||^2_{\mathcal{H}}\}$ (in case a)) or $\omega_\mu(W(F)) = \langle\Omega^{\mathcal{F}}, \rho_\mu(F)\Omega^{\mathcal{F}}\rangle$ (in case b)).

$\hat{\varphi}(F) := a^*(KF) + a(KF)$ is the usual field operator on $\mathcal{F}^s(\mathcal{H})$ and we can determine the ("symplectically smeared") two-point function as

$$\begin{aligned} \lambda^{(2)}(F_1, F_2) &= \langle\Omega^{\mathcal{F}}, \hat{\varphi}(F_1)\hat{\varphi}(F_2)\Omega^{\mathcal{F}}\rangle \\ &= \langle KF_1, KF_2\rangle_{\mathcal{H}} \\ &= \mu(F_1, F_2) + \frac{i}{2}\sigma(F_1, F_2) \end{aligned} \tag{3.82}$$

for $F_1, F_2 \in \Gamma$, resp. the "four-smeared" (Wightman) two-point distribution as

$$\Lambda^{(2)}(f_1, f_2) = \lambda^{(2)}\lambda(\begin{pmatrix}\rho_o E f_1 \\ \rho_1 E f_1\end{pmatrix}, \begin{pmatrix}\rho_o E f_2 \\ \rho_1 E f_2\end{pmatrix}\varrho) \tag{3.83}$$

for $f_1, f_2 \in \mathcal{D}(\mathcal{M})$. The fact that the antisymmetric (= imaginary) part of $\lambda^{(2)}$ is the symplectic form σ implies for $\Lambda^{(2)}$:

$$\begin{aligned} \mathrm{Im}\Lambda^{(2)}(f_1, f_2) &= -\frac{1}{2}\int_\Sigma [f_1 E' \rho_o' \rho_1 E f_2 - f_1 E' \rho_1' \rho_o E f_2]\, d^3\sigma \\ &= \frac{1}{2}\langle f_1, E f_2\rangle \end{aligned} \tag{3.84}$$

by eqn. (3.64). All the other n-point functions can also be calculated; one finds that they vanish if n is odd and that the n-point functions for n even are sums of products of two-point functions. For example, the four point function is given by:

$$\lambda^{(4)}(F_1, F_2, F_3, F_4) = \lambda^{(2)}(F_1, F_2)\lambda^{(2)}(F_3, F_4) + \lambda^{(2)}(F_1, F_3)\lambda^{(2)}(F_2, F_4) + \lambda^{(2)}(F_1, F_4)\lambda^{(2)}(F_2, F_3) \tag{3.85}$$

This restriction to quasifree states is not a 'natural' one; rather they are chosen for their mathematical simplicity (they are exclusively determined by their two-point function) and by the fact that we are dealing with linear fields.

3.4.2 Hadamard states

We now specialise further, i.e. make greater restrictions on the notion of state. This leads us to the notion of Hadamard states. First some geometrical concepts are required[30]:

Definition 3.7 *Let Σ be a spacelike Cauchy surface of $(\mathcal{M}, g)$.*
A **causal normal neighborhood** *N of Σ is an open neighborhood of Σ in $\mathcal{M}$ such that Σ is a Cauchy surface for N and such that for all $x_1, x_2 \in N$ with $x_1 \in J^+(x_2)$ there exists a convex normal neighborhood which contains $J^-(x_1) \cap J^+(x_2)$. (As a consequence, the squared geodesic distance $\sigma(x_1, x_2)$ is then well defined and smooth for all causally related pairs of points in N).*

Lemma 3.8 (Lemma 2.2 of [31]) *For each spacelike Cauchy surface Σ there exists a causal normal neighborhood N.*

We choose a preferred time orientation on $(\mathcal{M}, g)$ and a smooth global time function $T : \mathcal{M} \to \mathbf{R}$ which increases towards the future. Let $\mathcal{O} \subset \mathcal{M} \times \mathcal{M}$ be an open neighborhood of the set of causally related points (x_1, x_2) such that $J^+(x_1) \cap J^-(x_2)$ and $J^-(x_1) \cap J^+(x_2)$ are contained within a convex normal neighborhood and $\mathcal{O}'$ an open neighborhood in $N \times N$ of the set of causally related points such that $\overline{\mathcal{O}'} \subset \mathcal{O}$.
Within $\mathcal{O}$ the squared geodesic distance $\sigma(x_1, x_2)$ is well defined and we define for each $n \in \mathbf{N}$ a real function $v^{(n)} \in \mathcal{C}^\infty(\mathcal{O})$ as the power series

$$v^{(n)}(x_1, x_2) := \sum_{m=0}^{n} v_m(x_1, x_2)\sigma^m \tag{3.86}$$

where the v_m are uniquely determined by the Hadamard recursion relations [21].
Let $\chi \in \mathcal{C}^\infty(N \times N)$ be a function with the property that

$$\chi(x_1, x_2) = \lambda\{ \begin{array}{ll} 0, & \text{for } (x_1, x_2) \notin \mathcal{O} \\ 1, & \text{for } (x_1, x_2) \in \mathcal{O}'. \end{array} \varrho.$$

For each $n \in \mathbf{N}$ and $\epsilon > 0$ we define in $\mathcal{O}$ the (complex valued) function

$$\begin{aligned} G_\epsilon^{T,n}(x_1, x_2) := \frac{1}{(2\pi)^2}\lambda\Big(&\frac{\Delta(x_1, x_2)^{1/2}}{\sigma + 2i\epsilon(T(x_1) - T(x_2)) + \epsilon^2} \\ &+ v^{(n)}(x_1, x_2)\ln(\sigma + 2i\epsilon(T(x_1) - T(x_2)) + \epsilon^2)\varrho\Big), \end{aligned} \tag{3.87}$$

where the branch-cut for the logarithm is taken to lie along the negative real axis. This brings us to the vital definition::

Definition 3.9 *Let $(\mathcal{M}, g)$ be a globally hyperbolic manifold, Σ a Cauchy surface of $\mathcal{M}$, N a causal normal neighborhood of Σ and $\chi, T, G_\epsilon^{T,n}$ as above. Then we call a quasifree state ω of the Weyl-algebra $\mathcal{A}$ of the Klein-Gordon field on $(\mathcal{M}, g)$ a* **(global) Hadamard state** *if its two-point distribution $\Lambda^{(2)}$ is such that there exists a sequence of functions $H^n \in \mathcal{C}^n(N \times N)$ such that for all $f_1, f_2 \in \mathcal{C}_o^\infty(N)$ and all $n \in \mathbf{N}$ we have*

$$\begin{aligned} \Lambda^{(2)}(f_1, f_2) &= \lim_{\epsilon \to 0} \int_{N \times N} \Lambda_\epsilon^{T,n}(x_1, x_2) f_1(x_1) f_2(x_2)\, d^4\mu(x_1) d^4\mu(x_2) \\ \text{where } \Lambda_\epsilon^{T,n}(x_1, x_2) &:= \chi(x_1, x_2) G_\epsilon^{T,n}(x_1, x_2) + H^n(x_1, x_2). \end{aligned} \tag{3.88}$$

Note that χ was chosen to be zero where $G_\epsilon^{T,n}$ was not defined, so $\Lambda_\epsilon^{T,n}$ is well defined throughout $N \times N$. Kay and Wald [31] show that the definition is actually independent of the choice of N, χ and T. The Definition 3.9 above is independent of the choice of Σ, too.

3.4.3 Finally: Renormalisation

Probably the most efficient regularisation technique, and the one favoured by most relativists working on quantum fields in curved spacetimes, is that of 'point splitting'. Thus method is best illustrated by reexamining the notion of normal ordering in Minkowski spacetime. Recall that the physics behind this scheme was the subtraction of the (unobservable and infinite) vacuum stress energy to get coherent results. Here we shall exhibit exactly how this is implemented in a rigourous sense.

First of all, we must realise that the essential difficulty with the stress tensor is that it involves expressions of the form $< \varphi^2 >$, which means that we are trying to calculate the product of two distributions at the same spacetime point. This is a non-convergent procedure, as we have stated before. However, the quantity $< \varphi(x)\varphi(x') >$ is a more reasonable object; it is a bi-distribution. For 'physically reasonable' states (in some Fock space) the singular behaviour of this bi-distribution as $x' \to x$ will be the same as for the vacuum expectation value, $< 0|\varphi(x)\varphi(x')|0 >$. So for such states the difference

$$F(x,x') \equiv < \varphi(x)\varphi(x') > - < 0|\varphi(x)\varphi(x')|0 > \tag{3.89}$$

will be a smooth function of x and x' (and thus not a distribution!). There is no difficulty with taking coincidences in such functions, of course; then we can *define* the pathological object $< \varphi^2(x) >$ as

$$< \varphi^2(x) > = \lim_{x' \to x} F(x,x') \tag{3.90}$$

Similarly

$$< T_{\mu\nu}(x) > = \lim_{x \to x'} \{\nabla_\mu \nabla'_\nu F(x,x') - \frac{1}{2} g_{\mu\nu}[(\nabla_\varepsilon \nabla'^\varepsilon + m^2)F(x,x')]\} \tag{3.91}$$

Point splitting is equivalent to normal ordering in Minkowski spacetime. The subtraction procedure is equivalent to the renormalisation of constants in the generalised Einstein action which we considered earlier ([?]).

However, in a general curved spacetime, we have no recourse to a natural vacuum state. Thus the normal ordering procedure of 'subtracting off the vacuum energy' is not as easy as it is in the case of Minkowski space. [However, we note here that differences in the expected stress energy between two states still can be defined by the procedure above i.e. $< T_{\mu\nu} >_1 - < T_{\mu\nu} >_2$ is well defined for any two algebraic states ω_1, ω_2 which possess a two-point function (see eqn.(3.78) and for which $< \varphi(x)\varphi(x') >_1 - < \varphi(x)\varphi(x') >_2$ is a smooth function.]

How are we to proceed then in curved spacetime? It turns out that the best way is via an axiomatic scheme proposed by Wald [50], i.e. we postulate properties we would expect a reasonable $< T_{\mu\nu} >$ to have:

1. The natural, well defined expression for the difference in expected stress energy between two states (given by the above prescription) should be valid.

2. The expected stress-energy should be local with respect to the state of the field.(see [50] for precise details)

3. For all states, we have $\nabla^\mu < T_{\mu\nu} >= 0$

4. In Minkowski spacetime, we have $< 0|T_{\mu\nu}|0 >= 0$

Note that these axioms refer to a physically plausible (finite, possible renormalised) stress- tensor;if we can obtain such an object, then it will be unique up to the addition of local curvature terms.[2] But we are still some way off actually exhibiting that such an object exists. To construct our desired tensor,

[2]See discussion on renormalisation of Newton's constant, G_B, in section 3.3.

we turn to the point-splitting procedure in curved spacetime. This technique starts with the Hadamard elementary solution

$$H(x,x') = \frac{(\Delta)^{\frac{1}{2}}}{(2\pi)^2[\sigma + 2i\epsilon(t-t') + \epsilon^2]}(x,x') + V(x,x')ln[\sigma + 2i\epsilon(t-t') + \epsilon^2] + W(x,x') \tag{3.92}$$

with

$$V(x,x') = \Sigma_{j=0}^{\infty} v_j(x,x')\sigma^j \tag{3.93}$$

$$W(x,x') = \Sigma_{j=0}^{\infty} w_j(x,x')\sigma^j \tag{3.94}$$

The potential problems in the convergence of these sums is circumvented in what we are about to do. First, we may solve for v_i and $\Delta(x,x')$ along the geodesic joining x and x' by substituting in the Klein-Gordon equation [21]. As for the w_j, if we specify w_0, then all w_j for $j > 0$ are uniquely determined. In Minkowski spacetime, $w_0 = 0$; this choice ensures that $H(x,x')$ has the same singularity structure as $< 0|\varphi(x)\varphi(x')|0 >$. We retain $w_0 = 0$ in the curved space case, so that the short-distance singularity structure of $H(x,x')$ is as close as possible to that of the Minkowski two-point distribution.

With this elementary solution in hand, we proceed to construct $< T_{\mu\nu} >$ in some state, ω, as follows:
(i) calculate $G(x,x') =< \varphi(x)\varphi(x') >$[3]
(ii) form $F(x,x') = G(x,x') - H(x,x')$
(iii) define $< T_{\mu\nu} >= \lim_{x' \to x}\{\nabla_a \nabla'_b F(x,x') - \frac{1}{2}g_{ab}[\{\nabla_c \nabla'^c + m^2\}F(x,x')]$

Now it turns out that the Hadamard states have the necessary properties we seek to define a meaningful stress tensor. We require that $G(x,x')$ above is the two-point distribution of a Hadamard state as in Definition 3.9. It is easy to see that all the singular terms in $G(x,x')$ will be cancelled by the subtraction of $H(x,x')$. It must be stressed that *all* Hadamard states have the same singularity structure (in a given spacetime). The information which distinguishes different Hadamard distributions is contained in the function $W(x,x')$.

[3]See remarks accompanying eqn 3.78.

As mentioned before, the subtraction in (ii) is identical to the subtraction performed in section 3.3 in the formation of L_{ren} as $L_{ren} = L_{eff} - L_{div}$. With this subtraction, we also solve several problems:

- Only the $j = 0, 1$ terms survive in the series for V and W, thus avoiding questions of convergence of the associated sums.

- The prescription for $< T_{\mu\nu} >$ satisfies all of Wald's axioms [8]

Instead of subtracting $H(x, x')$ in the above, we could also make use the DeWitt- Schwinger expansion of the Feynman Green function which we encountered earlier (see section 3.3).

How does the point splitting method of obtaining $< T_{\mu\nu} >$ compare with the approach based on the effective action?

- We have already noted the computational advantages of the former: in a sense all we need to know are the elementary operations of subtraction and differentiation, while for the effective action method to work, we would require a highly non-trivial functional differentiation (with respect to $g_{\mu\nu}$) to be performed.

- On the other hand, the similarities between the two methods stem from their physical motivations: recall than in the case of L_{eff}, we obtained the divergent terms as arising from the short wavelength behaviour of the field; these were then absorbed into the gravitational part of the Lagrangian. The point splitting method also recognises the dangerous nature of the high frequency modes: we desired to obtain an $H(x, x')$ which reduced to $< 0|\varphi(x)\varphi(x')|0 >$ in the case of Minkowski spacetime (as in eqn.(3.89)). Thus in curved spacetime, $H(x, x')$ will model quite closely the behaviour of the Minkowski two-point distribution.

And that is exactly what we want: we obtain a smooth $F(x, x')$ in (ii) above if $G(x, x')$ has similar short-distance-singularity structure to the Minkowski vacuum. This is reasonable and thus we are able to obtain a $< T_{\mu\nu} >$. (In fact, $< T_{\mu\nu} >$ is well-defined and nonsingular for all Hadamard states.) So then the prescription of subtraction removes all the singular terms.

So a necessary condition for a state, ω, to be physically acceptable is that it be a Hadamard state. For this to be of any value for us, we certainly require that there should be 'enough' of them in a suitable sense. It is known that in any globally hyperbolic spacetime there is always a class of quantum states, forming a dense subspace of a Hilbert space, whose two-point functions have the Hadamard singularity structure (3.88) (see [19]) .
This is based on the result that in static, globally hyperbolic spacetimes, the static vacuum state is a Hadamard state. Also, if a state, ω, satisfies the Hadamard condition in a causal normal neighbourhood of any Cauchy surface, then it satisfies that condition throughout the spacetime. In the Fock space associated with this vacuum state, a dense set of vectors satisfy the Hadamard condition. Then the existence of a large class of Hadamard states in an arbitrary globally hyperbolic spacetime follows by performing an appropriate deformation [54].

This completes our discussion on the renormalisation of the stress-tensor using the point-splitting approach. But the general question of the future role of $\langle T_{\mu\nu} \rangle$ remains unanswered (and perhaps even unsaid); nonetheless is clear that the Hadamard states will have an important part to play in this. Thus barring any major revisions, the folium of states generated by the Hadamard states is a good candidate for the set of physical states of scalar quantum fields on globally hyperbolic spacetimes.

So we see that the semiclasical Einstein equation not only provides a means of determining the effect of quantum fields on (unquantised) spacetimes, it also restricts which quantum states are physically acceptable. Seen in this light, we have here another exhibition of the effect of geometry on quantum systems. One could attempt to counter this statement by claiming that in a quantum gravity scenario, all quantum states should be present (merely on grounds of not restricting anything). Being that as it may, at least in the 'semiclassical' case we have examined, the above remarks remain valid. Besides, the conditions under which we would expect quantum gravity to apply, viz cases of extreme curvature, might even result in a further restriction on the allowable quantum states. Ths status of these ideas is not clear at present.

3.5 Back to Berry!

We now attempt to justify our foray into the two apparently disparate fields we have exhibited above. The connection between the Berry phase and the black hole scenario arises from the realisation that in appropriate circumstances, the treatment of the semiclassical system of gravity and matter may be accomplished using the analysis of the Born-Oppenheimer we have outlined in section (1.6). Before delving into the details of what this entails, we provide some motivation.

Remarks:

- Perhaps we are being naive in attempting to draw together two apparently distant topics using what is after all a method first developed in quantum chemistry [28].
- But in mitigation, one would expect that the evaporation of black holes and subsequent backreaction is in some sense a *slow* process, at least until the black hole reaches its final stages. This harks back to Berry's orginal derivation[9]. So while this link may be somewhat tenuous, it is nevertheless worthy of further study.

The idea of dividing the system into a (classical) gravitational part and a (quantized) matter part is a natural one in light of the absence of a fully-fledged quantum theory of gravity. This means that the two sub-systems are governed by coupled equations:

$$i\hbar\frac{\partial}{\partial t}\psi(\phi,g) = \hat{H}\psi(\phi,g) \tag{3.95}$$

$$G_{ij} = 8\pi M_P^{-2} < \psi|T_{ij}|\psi > \tag{3.96}$$

It is possible to obtain the above from the Wheeler-deWitt equation using an approximation scheme where the source-free equation $G_{ij} = 0$ arises at the leading order (in the Planck mass). The next order M_P^0 produces eqn.(3.95) above. Now the crux of the matter [11] is that the stress-energy expectation value $< \psi|T_{ij}|\psi >$ is determined by a typical phase of the matter wave function $\psi(\phi, g)$ This phase is a perturbation on the leading order phase (for

$G_{ij} = 0$). We require that the dispersion in the metric derivative of the matter phase should be negligible. This condition is satisfied by the class of adibatically varying metrics.

Using the ideas of the Berry phase, we may arrive at the above considerations with more rigour. The idea is to employ a toy model which mimics the form of a minisuperspace, with a 'heavy' particle, Q, representing gravity and a 'light' particle, q, representing the matter. The matter Hamiltonian has the form

$$h(Q) = \frac{1}{2}m\dot{q}^2 + u(q, Q) \tag{3.97}$$

Then using the Born-Oppenheimer approximation, the effective Hamiltonian for the Q mode turns out to be

$$H_{eff} = -\frac{\hbar}{2M}(\nabla_Q - iA(Q))^2 + MV(Q) + \epsilon_n Q \tag{3.98}$$

where $\epsilon_n(Q)$ is the eigenvalue for the matter Hamiltonian. From the above, we see that the Q mode is 'disturbed' by the q mode, with the effective momentum of the former given by

$$P = P_0 + \hbar A \tag{3.99}$$

where P_0 is the source-free Q-momentum. This last equation in fact determines the backreaction in eqn. 3.95 above (in the semiclassical limit when $M \to \infty$). Using an expansion of the total system wavefunction, $\Phi(Q, q) = e^{\frac{iS(Q,q)}{\hbar}}$ wherein

$$S(Q, q) = MS_0(Q, q) + S_1(Q, q) + M^{-1}S_2(Q, q) + \ldots \tag{3.100}$$

we arrive at the 'leading order' equation

$$\frac{1}{2M}P_0^2 + MV(Q) = E, \qquad P_0 = M\nabla_Q S_0 \tag{3.101}$$

This is the Hamilton-Jacobi equation for the source-free Q mode. The evolution of the q mode comes about from the next-to-leading order equation:

$$i\hbar\frac{\partial}{\partial t}n(q, Q) = h(Q)n(q, Q) \tag{3.102}$$

The Berry connection has the form

$$A(Q) = -(\hbar \nabla_Q S_0)^{-1} \langle n|h|n\rangle \tag{3.103}$$

From this follows the backreaction equation

$$\frac{P^2}{2M} + MV(Q) + \langle n|h|n\rangle = \epsilon \tag{3.104}$$

when we substitute from eqn. 3.101 into eqn. 3.99 for P_0. So in the adiabatic limit, the semiclassical backreaction gets completely determined by the Berry connection; a rather remarkable suggestion! We have come a long way indeed from the study of Berry's connection in its original context. And judging by these results, there is still plenty of scope for further developments.

Acknowledgment

I would like to thank my supervisor, Professor George Ellis, whose original comments on the Berry phase motivated this study. For financial support, I am indebted to the Foundation for Research and Development (FRD) for a bursary.

In the course of any work conducted over a lengthy period of time, one accumalates certain debts, both emotional and academic. For their assistance, I would like to thank the following:
Profs. Paul Cherenack and Neven Bilic for stimulating courses; the latter also for useful comments on the manuscript. In this regard, Dave Hislop deserves an honourable mention for a thorough check; also to Tim Gebbie. Then to my office mates, erstwhile and otherwise, Tim, Conrad and Nico, thanks are due for numerous discussions on physics as well as the view from the window. To Marty for putting up with it all and for the squash (not forgetting the Moose here, of course). And to the semi-permanent occupants of Room 401, especially André, thanks for the tips on the machine. To Prof. Malcolm MacCallum, for a timely intervention on the part of gravity. It goes almost without saying that my family have been very encouraging and tolerant of a sometimes wayward son and sibling.

To Ginny, without whose support I could not function.

Appendix A

Notation and conventions

supp f: the support of the function f, i.e. the closure of the set of points at which f does not vanish
X an open subset of $\mathbb{R}^n$
$\mathcal{E}(X) \equiv \mathcal{C}^\infty(X)$: space of smooth (infinitely differentiable) functions $f : X \to \mathbb{C}$
$\mathcal{E}'(X)$: dual space of $\mathcal{E}(X)$. It is the space of distributions with compact support.
$\mathcal{D}(X) \equiv \mathcal{C}_o^\infty(X) = \{f \in \mathcal{C}^\infty(X);\ supp\, f \text{ compact}\}$: space of testfunctions with compact support.
$\mathcal{D}'(X)$: dual space of $\mathcal{D}(X)$, the space of distributions in X
$\mathcal{S}(\mathbb{R})$: Schwartz space of functions in $\mathcal{C}^\infty(\mathbb{R})$ that are rapidly decaying
$\mathcal{S}'(\mathbf{R}^n)$: the dual space of $\mathcal{S}(\mathbf{R}^n)$. It is the space of tempered distributions.
Since $\Delta\mathbf{R}^n \subset \mathcal{S}(\mathbf{R}^n) \subset \mathcal{E}(\mathbf{R}^n)$ we have $\mathcal{E}'(\mathbf{R}^n) \subset \mathcal{S}'(\mathbf{R}^n) \subset \mathcal{D}'(\mathbf{R}^n)$.

Aspacetime$(\mathcal{M}, g)$ is time-orientable if there exists a smooth timelike vector field on $\mathcal{M}$, *i.e.* we can unambiguously distinguish the future light cone from the past light cone throughout the manifold.
$(\mathcal{M}, g)$ is globally hyperbolic if it possesses a Cauchy surface Σ, *i.e.* a 3-dim. hypersurface that is intersected by each inextendible causal (null or timelike) curve in $\mathcal{M}$ exactly once. By a theorem due to Geroch[54], $(\mathcal{M}, g)$ may be foliated as $\mathcal{M} = \mathbb{R} \times \Sigma$, where the hypersurface Σ is a spacelike Cauchy surface for $(\mathcal{M}, g)$.

If S is a subset of $\mathcal{M}$, we define $J^+(S)$ (resp. $J^-(S)$) to be the set of all points $x \in \mathcal{M}$ such that x can be connected to a point in S by a future-directed (resp. past-directed) causal curve from S to x in $\mathcal{M}$.A convex normal neighborhood in $\mathcal{M}$ is an open set $U \subset \mathcal{M}$ such that for any two points $x_1, x_2 \in \mathcal{M}$ there exists a unique geodesic contained in U which connects x_1 and x_2.
$\sigma(x_1, x_2) = \pm\lambda(\int_a^b \lambda|g_{\mu\nu}(x(\tau))\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\varrho|^{1/2}\, d\tau\, \varrho)^2$ is the square of the geodesic distance from x_1 to x_2 in a convex normal neighborhood U, where $[a, b] \to \mathcal{M}, \tau \mapsto x(\tau)$ is the unique geodesic curve in U from x_1 to x_2 ($+$ ($-$) is chosen if $x(\tau)$ is spacelike (timelike)).

Appendix B

Algebraic Preliminaries

We collect together here some important notions and results concerning the algebraic approach to quantum field theory.

1. An algebra $\mathcal{A}$ over the field of complex numbers is a set that satisfies the following conditions.

   - $\mathcal{A}$ is closed with respect to a commutative $=$, associative, binary operation of addition, and forms a group w.r.t. this operation. The identity element of this group is termed the zero element, 0, of $\mathcal{A}$.
   - $\mathcal{A}$ is closed w.r.t. an associative, binary operation of multiplication, which is distributive w.r.t. addition. In cases where this operation is commutative, $\mathcal{A}$ is termed Abelian.
   - $\mathcal{A}$ is closed w.r.t multiplicaton by complex numbers, this operation satisfying the conditions
     $(\lambda_1 + \lambda_2)A = \lambda_1 A + \lambda_2 A;\ \lambda_1(\lambda_2 A) = (\lambda_1\lambda_2)A;$
     and
     $(\lambda_1 A_1)A_2 = A_1(\lambda A_2) = \lambda(A_1 A_2)$
     $\forall \lambda, \lambda_1, \lambda_2 \in \mathbb{C}$ and all $A, A_1, A_2 \in \mathcal{A}$

2. An algebra $\mathcal{A}$ is termed a $*$-algebra if it is equipped with a mapping $A \rightarrow A^*$ of $\mathcal{A}$ onto itself such that

- $(A^*)^* = A$; $(A_1 + A_2)^* = A_1^* + A_2^*$;
- $(A_1 A_2)^* = A_2^* A_1^*$;
- $(\lambda A)^* = \bar{\lambda} A^*$

Such a mapping is termed an involution.

3. A $*$-algebra $\mathcal{A}$ is said to be normed if it is equipped with a mapping $A \rightarrow \| A \|$ of $\mathcal{A}$ into the non-negative real numbers such that

   - $\| A_1 + A_2 \| \leq \| A_1 \| + \| A_2 \|$;
   - $\| A_1 A_2 \| \leq \| A_1 A_2 \|$;
   - $\| \lambda A \| = | \lambda | \| A \|$
   - $\| A \| = 0 \Leftrightarrow A = 0$.

   The mapping $\| \cdot \|$ is termed a norm. A normed algebra which is complete is called a Banach algebra.

4. A C^*-algebra is a normed $*$-algebra, $\mathcal{A}$, for which $\| A^* A \| = \| A \|^2$ and which is complete w.r.t that norm. Thus, if $\mathcal{A}_0$ is a normed $*$-algebra whose elements A satisfy the above requirement, then the algebra obtained by adding to $\mathcal{A}_0$ the limit points of its (normwise) Cauchy sequences is a C^*-algebra, $\mathcal{A}$, termed the norm-completion of $\mathcal{A}_0$.
   example: The set $\mathcal{L}(\mathcal{H})$ of all bounded operators in a HIlbert space $\mathcal{H}$ (with addition, multiplication and norm defined in the usual way for such operators) is a non-Abelian C^*-algebra, except when $\mathcal{H}$ is one-dimensional.

5. A representation, π, of $\mathcal{A}$ is a mapping of $\mathcal{A}$ into the bounded operators in a Hilbert space, $\mathcal{H}$, that preserves the *- algebraic structure, i.e.

   - $\pi(\lambda_1 A_1 + \lambda_2 A_2) = \lambda_1 \pi(A_2) + \lambda_2 \pi(A_2$;
   - $\pi(A_1 A_2) = \pi(A_1)\pi(A_2)$;
   - $\pi(A^*) = \pi(A)^*$

In general, $\| \pi(A) \| \leq \| A \|$, with equality holding $\forall A \in \mathcal{A} \Leftrightarrow \pi$ is faithful, i.e. $if \pi(A) = 0 \Rightarrow A = 0$.

6. The representation, π of $\mathcal{A}$ in $\mathcal{H}$ is termed *cyclic* if $\mathcal{H}$ contains a vector Φ such that $\pi(\mathcal{A})\Phi$ is dense in $\mathcal{H}$. Φ is then termed cyclic for π.

Bibliography

[1] Y. Aharanov and A. Anandan. Phase Change during a Cyclic Quantum Evolution *Phys. Rev. Lett.*, 58:1593–1596, 1987.

[2] Y. Aharanov, J. Anandan and L. Vaidman. Meaning of the wave function. *Phys. Rev. A*, 47:4616–4626, 1993.

[3] J. Anandan. Remarks concerning the geometries of gravity and gauge fields, in *Directions in General Relativity*, B.L. Hu, M.P. Ryan and C.V. Vishveshwara eds., Cambridge Univ. Press, Cambridge, 1993.

[4] J. Anandan. On the reality and geometry of quantum states. *gr-qc* 9505011.

[5] J. Anandan. Gravitational phase operator and cosmic strings. *Phys. Rev. D*, in press; *gr-qc* 9507049.

[6] A. Ashtekar and A. Magnon. Quantum fields in curved space-times. *Proc. Roy. Soc. A*, 346:375–394, 1975.

[7] A. Asorey, J.F. Cariñena and M. Paramio. Quantum evolution as a parallel transport. *J. Math. Phys.*, 23:1451–1458, 1982

[8] N.D. Birrell and P.C.W. Davies. *Quantum Fields in Curved Space.* Cambridge Univ. Press, Cambridge, 1982.

[9] M.V. Berry. Quantal Phase Factors Accompanying Adiabatic Changes. *Proc. Roy. Soc. Lond. A*, 392:45–57, 1984

[10] C. Crnkovic and E. Witten. Covariant description of canonical formalism in geometrical theories, in *300 Years of Gravitation*, S.W. Hawking, W. Israel, eds. Cambridge Univ. Press, Cambridge, 1987.

[11] D.P. Datta. Semiclassical backreaction and Berry's phase. *Modern Physics Letters A*, 8:191-196, 1993.

[12] P.C.W. Davies. Particles do not exist, in *Quantum Theory of Gravity*, S.M. Christensen, ed., Adam Hilger Press, Bristol, 1984.

[13] J. Dimock. Algebras of local observables on a manifold. *Comm. Math. Phys.*, 77:219–228, 1980.

[14] P.A.M. Dirac. Development of the physicist's conception of nature, in *The Physicist's Conception of Nature*, J.Mehra ed., D.Reidel, Dordrecht, 1973.

[15] T.S. Eliot. *Little Gidding.* Faber and Faber, London, 1942.

[16] C.J. Fewster and A. Higuchi. Quantum field theory on certain non-globally hyperbolic spacetimes. *preprint* BUTP-95/31, 1995.

[17] K. Fredenhagen and R. Haag. On the derivation of Hawking radiation associated with the formation of a black hole. *Comm. Math. Phys.*, 127:273–284, 1990.

[18] S.A. Fulling. What have we learned from quantum field theory in curved spacetime? in *Quantum Theory of Gravity*, S.M. Christensen, ed., Adam Hilger Press, Bristol, 1984.

[19] S.A. Fulling, F.J. Narcowich, and R.M. Wald. Singularity structure of the two-point function in quantum field theory in curved spacetime, II. *Ann. Phys. (NY)*, 136:243–272, 1981.

[20] S.A. Fulling, M. Sweeny, and R.M. Wald. Singularity structure of the two-point function in quantum field theory in curved spacetime. *Comm. Math. Phys.*, 63:257–264, 1978.

[21] P.R. Garabedian. *Partial Differential Equations.* John Wiley & Sons, New York, 1964.

[22] G. Giavarini, E. Gozzi, D. Rohrlich and W.D. Thacker. On removing Berry's phase. *Phys. Lett. A*, 138:235–241, 1989.

[23] D. Graudenz. On the spacetime geometry of quantum systems. *gr-qc* 9411.

[24] R. Haag and D. Kastler. An algebraic approach to quantum field theory. *J. Math. Phys.*, 5:848–861, 1964.

[25] S.W. Hawking. Particle creation by black holes. *Comm. Math. Phys.*, 43:199–220, 1975.

[26] S.W. Hawking. Zeta function regularisation of path integrals in curved spacetime. *Comm. Math. Phys.*, 56:133–162, 1977.

[27] C. Hong-Mo and T. Sheung Tseu. *Some Elementary Gauge Theory Concepts.* World Scientific, Singapore, 1993.

[28] R. Jackiw. Three Elaborations on Berry's connection, curvature and phase. *Int. J. Mod. Phys. A*, 3:285–297, 1988.

[29] J.M. Jauch. The Mathematical Structure of Elementary Quantum Mechanics. in *The Physicist's Conception of Nature*, J. Mehra ed., D. Reidel, Dordrecht, 1973.

[30] W.Junker. Adiabatic vacua and Hadamard states for scalar quantum fields on curved spacetime. *Ph.D. Thesis*, Univesität Hamburg, Hamburg, 1995.

[31] B.S. Kay and R.M. Wald. Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon. *Phys. Rep.*, 207:49–136, 1991.

[32] M. Nakahara. *Geometry, Topology and Physics.* Adam Hilger Press, Bristol, 1990.

[33] R. Penrose. Gravity and state vector reduction, in *Quantum concepts in space and time*, R.Penrose, C.J. Isham, eds., Clarendon Press, Oxford, 1986.

[34] E.Prugovečki. *Quantum Mechanics in Hilbert Space.* Academic Press, New York, 1971.

[35] M. Reed and B. Simon. *Methods of Modern Mathematical Physics II.* Academic Press Inc. London, 1975.

[36] C. Rovelli. What is observable in classical and quantum gravity? Class. & Quant. Grav., 8:297-316, 1991.

[37] J. Samuel and R. Bhandari. General Setting for Berry's Phase. Phys. Rev. Lett., 60:2339–2342, 1988.

[38] L.I. Schiff. *Quantum Mechanics.* McGraw-Hill, New York, 1955.

[39] R.N. Sen. Quantum field theories on Hilbert bundles. *Nuc. Phys. B (Proc. Suppl.)*, 6:167–169, 1989.

[40] A. Shapere and F. Wilczek. *Geometric Phases in Physics.* World Scientific, Singapore, 1989.

[41] B. Simon. Holonomy, the Quantum Adiabatic Theorem, and Berry's Phase. *Phys. Rev. Lett.*, 51:2167–2170, 1983.

[42] L. Sriramkumar. Limits on the validity of the semiclassical theory. *preprint* IUCAA-15/95, 1995.

[43] R.F. Streater and A.S. Wightman. *PCT, Spin and Statistics, and All That.* W.A. Benjamin, New York, 1964.

[44] A. Trautman. Fibre Bundles, gauge fields and gravitation, in *General Relativity and Gravitation 1*, A. Held, ed., Plenum Press, New York, 1979.

[45] W.G. Unruh. Particles and Fields, in *Quantum Mechanics in Curved spacetime.* J.Audretsch and V. DeSabbata, eds. Plenum Press, New York. 1990.

[46] M. Veltman. *Diagrammatica: The path to Feynman diagrams.* Cambridge Univ. Press, Cambridge, 1994.

[47] J. Vinet. Invariant Berry connections. *Phys. Rev. D.*, 37:2369–2372, 1988.

[48] R.M. Wald. On particle creation by black holes. *Comm. Math. Phys.*, 45:9–34, 1977.

[49] R.M. Wald. The back reaction effect in particle creation in curved spacetime. *Comm. Math. Phys.*, 54:1–19, 1977.

[50] R.M. Wald. Trace anomaly of a conformally invariant quantum field in curved spacetime. *Phys. Rev. D*, 17:1477–1484, 1978.

[51] R.M. Wald. The existence of the S-matrix in quantum field theory in curved spacetime. *Comm. Math. Phys.*, 70:221–, 1979.

[52] R.M. Wald. Quantum gravity and time reversibility. *Phys. Rev. D*, 21:2742–2755, 1980.

[53] R.M. Wald. *General Relativity.* The University of Chicago Press, Chicago, 1984.

[54] R.M. Wald. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* The University of Chicago Press, Chicago, 1994.

[55] J.A. Wheeler and W.H. Zurek, eds. *Quantum Theory and Measurement* Princeton Univ. Press, Princeton, 1983.